

The doctrine of unripe time

Insistence on perfection is a well-known enemy of achievement, as is the belief that good works are best put off to the most favourable opportunity. Mr Caspar Weinberger knows how to play that game.

St Augustine is known, among other things, for having said "Lord make me chaste, but not yet". The governments involved in the endless and confusing negotiations on arms control seem in the same case. The chief negotiators, the United States and the Soviet Union, say that winning an agreement is the peak of their ambition — yet nothing tangible has been agreed in four years of talks. Much of the trouble is the augustinian doctrine of unripe time; this or that agreement might be fine, "but not yet".

Now the rot has spread to Western Europe, apparently discomfited by Mr Gorbachev's expressed willingness last week to fall in with Mrs Thatcher's earlier demand that there should be no agreement to get rid of nuclear missiles of intermediate range without an agreement on weapons with even shorter range. The obvious snag is that the defence of Western Europe is predicated on the possible use of nuclear weapons in repelling attack even by conventional forces. Thatcher's point in Moscow, that Soviet stocks of short-range nuclear weapons are even bigger than those held in the West, is only partly applicable; Soviet conventional forces are also bigger, which is why the time for an agreement is "not yet".

There is an outside chance that some of these difficulties will be resolved by the visit of Mr George Shultz, the US Secretary of State, to Moscow this week. For different reasons, the Soviet and US governments are both eager to strike a deal. Shultz's task is to find the basis of an agreement that could be reached before US politicians are immersed in the impending election campaign. He may even have been helped to that end by the remarkable article by his cabinet colleague, Mr Caspar Weinberger, in the *New York Times* at the weekend. The US Secretary of Defense set out to distinguish between "real" and other kinds of arms control agreements, the latter by definition unsatisfactory and even dangerous. Could it be that Weinberger fears that Shultz may return from Moscow with a piece of paper which is less than ideal in his eyes, and that some kind of public warning was called for? If so, Shultz may be impelled uncharacteristically to grab the bit between his teeth.

Weinberger's argument is familiar — that satisfactory negotiations can be conducted only from strength, and that the past few years have shown that US insistence on its strategic interests (the right to pursue the Strategic Defense Initiative or SDI, and the exclusion of British and French nuclear weapons from calculations of intermediate missile strengths) at Geneva has eventually won Soviet compliance. Technically, Weinberger is correct. SDI seems no longer a stumbling block to a deal on intermediate missiles, while the Soviet Union appears to accept that the British and French forces are a problem for the future.

On SDI, for example, it must be supposed that Soviet compliance with a continuing US research programme implies a Soviet decision to follow a similar path. Even though SDI will never be the all-protective defensive shield on which President Reagan has set his heart, the result will be that both adversaries will soon be equipped with weapons different in kind from any that now exist. Weinberger's successors may not then thank him for his insistence now that SDI should go ahead.

The obvious analogy is historical — with the period immediately after the Second World War when the United States briefly

pondered whether the new technology of nuclear weapons should be sanitized by a form of international control. In the event, the Baruch plan was never put forward as a formal basis for negotiation. In retrospect, it seems clear that, if it had been so canvassed, it would have been turned down by Stalin's Soviet Union, then well on the way to a replication of US technology. But if it had been put forward and accepted before any substantial nuclear stockpiles had accumulated, it is more than probable that the military technology of nuclear energy would have been less oppressive than it has become. To be sure, the intervening four decades would not have been free from trouble. No doubt there would also have been more tension in Europe. But there is a high chance that the especially haunting features of the present strategic balance would not have materialized. May not the US decision not to negotiate on the sanitization of space turn out to have similarly unpalatable consequences?

Nor is it self-evident, as Weinberger says, that "arms control and even arms reductions are not ends in themselves; they are a means for strengthening national security". He is right to say that people cannot defend themselves with pieces of paper. He is also right to insist that arms control agreements that cannot be verified may be dangerous by lulling one side into a false sense of security. But he is surely wrong to claim that the only kinds of agreements worth having are cast-iron copper-bottomed deals that never impede their parties' pursuit of national security. What this calculation most of all omits is that, while arms control is not an end in itself, it may be a means to an important end — that of the mutual understanding between natural adversaries which is the perennial threat to nations' peace of mind.

Weinberger, in the *New York Times*, also scorns the rash of arms control agreements of the 1970s, but that is altogether too facile a reading of the history of a decade, beginning with President Richard M. Nixon and ending with President Jimmy Carter, that might have led somewhere if it had not been for Carter's political weakness and the Soviet adventure in Afghanistan. What that rash of agreements showed is that each superpower had a sufficient understanding of the other to anticipate its anxieties on security.

How, then, should one judge whatever Shultz brings back from Moscow? *Pace* Weinberger, the best achievement would be something likely to lead to something else. With the intermediate-range agreement so near completion at Geneva, but latterly cast in doubt by Thatcher's proper worry about short-range missiles, the best prize would be an agreement to cut Euromissiles by a half immediately, and to reduce them to zero in a reasonable period of time, say three years from now, when the short-range issue has been discussed. The benefit of such an agreement is that it would make the more difficult question of strategic arms more accessible, both in respect of the numbers in the arsenals of the two sides and of the independent British and French forces. But there is no chance, within the lifetime of the present US administration, of reaching the point at which President Carter finished — serious negotiations about a comprehensive test-ban treaty. Similarly, Gorbachev's dream of returning to a world free from nuclear weapons of any kind also belongs to the augustinian not now. □



Pentagon shoots across space station's bows

- Internal dispute threatens progress
- Military options cause global row

Washington

THE issue of whether the US manned space station will be built with the collaboration of the advertised partners overseas — Canada, the European Space Agency (ESA) and Japan — is coming rapidly to a head. It is now up to President Ronald Reagan, who must soon settle the dispute among several of his government agencies that could spell the end to international cooperation on the US manned space station.

The question is whether the Department of Defense can or should have unlimited access to the space station, something that other nations participating in the project have said they cannot accept.

The National Aeronautics and Space Administration (NASA), and the Departments of State, Defense and Commerce are all involved in the debate over international participation in the space station. The Defense Department showed little interest in the space station until last December. But late last year, the Pentagon asked that negotiations with other countries should be put on ice until its own role had been more fully articulated.

When negotiations resumed in February, it became clear that direct Pentagon involvement in the project would cause

serious problems for international partners, notably Japan and the members of ESA (the European Space Agency).

Government agencies have not so far been able to respond to these anxieties. Discussions were continued throughout March and into April, but ultimately broke down without an agreement. Each US agency involved has now prepared a briefing paper that will be delivered to President Reagan this week.

Meanwhile, the Defense Department has made its position public by releasing a letter dated 7 April, 1987, from Defense Secretary Caspar Weinberger to Secretary of State George Shultz. Weinberger warns of "paying too high a price for international cooperation". He urges that the US negotiating position should explicitly reserve the right to "conduct national security activities on the US elements of the Space Station, without the approval or review of other nations".

Weinberger is also anxious about a one-way flow of US space technology to international partners who could also be industrial competitors, and discouraging of the temptation to elevate the concept of "equal partnership" to the point at which it might dilute the symbol of US leadership in the space station.

NASA and the State Department have argued that any agency of the US government should be permitted to use the space station, provided that is consistent with international agreements on the peaceful use of space. NASA has consistently said that the space station would be used only for peaceful purposes, a position that may be at odds with Pentagon interests.

Ironically, the Defense Department has no specific plans for using the station, but is merely seeking to retain its options for the future. John Logsdon, an analyst of US space policy at George Washington University, says the current debate goes to the heart of future US activities in space. Logsdon says Weinberger's letter is an example of the Defense Department's fight for the "soul" of the US space programme.

Although Reagan has publicly endorsed international cooperation on a civil space station, he cannot afford to ignore the Pentagon's wishes. Reagan is being asked to decide the argument about policy on the space station at great speed: all parties are interested in resuming negotiations by the end of this month.

Joseph Palca

Another shock for Europe's space planes

London

CONTINUING pressure from the US Department of Defense to retain the option for a military use of the space station has surprised and disturbed the European Space Agency (ESA), says Dr Wilhelm Brado, head of cabinet at ESA in Paris.

The ESA covenant supports space research for exclusively peaceful purposes; the letter from the US Secretary of Defense, Mr Caspar Weinberger, to the Secretary of State, Mr George Shultz (see left), is clearly perceived as a threat.

Although the official US position has not been made explicit, European space administrators are clearly concerned at the implications of Pentagon pressure. "If they take this decision... we are completely out", declared Brado bluntly.

"This was meant to be a civilian space station", he said. "Is that offer being maintained, or is it being changed, with all the consequences that may entail?"

ESA's participation in the manned space station has been fraught with difficulties even without proposals for military use of the project. Negotiations on terms for the international effort have stumbled in the past over questions of financial returns and European autonomy.

The next negotiation between the US National Aeronautics and Space Administration and ESA had been scheduled for next week, but Brado says it is not now clear whether that meeting will go ahead. "Perhaps our delegation needs more time, or perhaps our American friends need more time to make a decision on the content of this letter before we can meet." He says the United States must decide if it will "go forward alone if the cost of cooperation is too high". He added: "We would certainly regret this decision — but if it is a clear-cut decision, then we would have to make our own decision."

ESA members agreed in 1985, that in the long term Europe should be autonomous in its space research, with the capability to launch payloads into orbit when and where it chooses. That plan may become more immediate. "Our ministers will have to reconsider the issue; it's not impossible that we'd decide to go completely independent", Brado says.

The first indications of a change in the US position came at the end of December, when the Department of Defense announced its intention to play a part in the space station. But Weinberger's letter setting out the Pentagon position underlines the difficulties in achieving international cooperation.

Kathy Johnston

Espionage denied

London

SOVIET protests to the French Foreign Ministry, refuting claims that Soviet citizens had attempted to obtain classified information on the Ariane rocket could "tarnish mutually beneficial cooperation between the two nations, particularly in the area of peaceful space research", according to Moscow radio. Commenting on the protests, Moscow's Francophone service for France and Belgium said the two countries had had a successful joint manned space mission in 1982, and that preparations are now under way for a second such mission.

Soviet denials of the alleged espionage have put special emphasis on the superiority of Soviet space technology. Ariane's cryogenic engine (the alleged target of espionage) will be able to put a payload of 21 tonnes into orbit not earlier than 1995, the Soviet commentators note, whereas the Soviet Proton carrier, which has a similar payload, has been in operation for more than five years.

Vera Rich

Soviet Union frozen out of Ocean Drilling Program partnership

Washington

PLANS to make the Soviet Union a full partner in the Ocean Drilling Program (ODP), the next phase of the exploration of the Earth's oceanic crust, may collapse in the face of objections from the US military, to the embarrassment of the sponsoring US agencies, chiefly the National Science Foundation (NSF).

Just one day before a US delegation was to set off for Moscow to sign an agreement making the Soviet Union a full partner in the Ocean Drilling Program, it was told (on 6 February) that it should not leave. The delay was supposed to be temporary, two weeks at most, pending resolution of national security issues raised by the Department of Defense (DoD).

But in the intervening two months, DoD's position has hardened. Citing worries about national security, DoD is now adamant that Soviet scientists should not be permitted access to the ODP research ship. Once regarded as a small step toward improved relations, Soviet participation in ODP has now become a political hot potato that has reached the very highest levels of US government.

Although the Soviet Union had been an active participant in the Deep Sea Drilling Program, ODP's predecessor, it was not invited to join ODP until 1985. The invitation followed a review by the National Oceanic and Atmospheric Administration, the State Department, the Navy and the interagency Committee on Exchanges (COMEX). The Navy's approval did mention potential difficulties in the area of technology transfer. Late last year the Soviet Union indicated its willingness to join ODP at the beginning of 1987.

But in November, the White House Office of Science and Technology Policy asked for a further review of Soviet participation. The investigation led to some minor revisions in the memorandum of understanding, but raised no serious objections. By then the date for Soviet membership had slipped to 1 March.

The NSF is the lead agency for ODP, providing approximately \$19 million towards the annual budget. Six international partners — the United Kingdom, West Germany, France, Canada, Japan and the European Science Foundation — each provide \$2.5 million to participate in

The Soviets could drill through the sea-bed, and bug our embassies FROM UNDERNEATH!



the programme. ODP members contribute scientists to each cruise, and help plan future drilling legs.

Everything seemed in order as a delegation from the State Department, NSF and the Joint Oceanographic Institutions prepared to go to Moscow in February for the signing ceremony. But only days before the ceremony, a letter from the assistant secretary of state for oceans and environmental and scientific affairs John

Negroponte to Navy deputy counsel Hugh O'Neill mentioning that the signing was about to take place set off alarms at DoD. Such was its concern that Secretary of Defense Caspar Weinberger was prevailed upon to ask President Reagan's national security adviser, Frank Carlucci, to intercede and stop the signing.

Both the Navy and the Office of the Secretary of Defense have concluded that giving Soviet scientists access to the technology aboard the ODP drilling ship would be a threat to national security. NSF assistant director William Merrell finds this argument far-fetched. NSF says there are only nine pieces of equipment on board the drilling ship that would require export licences, three of which will soon be removed from the control list.

Merrell says that there could be safeguards on board the ships to prevent unauthorized access if DoD insisted that certain equipment be kept off-limits to Soviet scientists. But DoD is not convinced that such measures would be adequate.

The National Security Council has now asked NSF, DoD and other interested agencies to prepare arguments on whether Soviet participation in ODP should be allowed to proceed. At this point, it would be acutely embarrassing to have to retract an invitation already accepted.

NSF argues that there are good scientific and political reasons to include Soviet scientists in ODP. For instance, NSF says that Soviet participation would help gain permission to drill in areas now out of bounds to ODP. Exploration of the Arctic seabed will be much easier with Soviet help, while Soviet survey ships have already studied the southern Indian Ocean where ODP hopes to go late this year. NSF has also raised the possibility that, if the Soviet Union is denied membership in ODP, it will start its own ocean drilling project and will thus become a competitor for international participants.

NSF is also concerned that rejecting Soviet membership would send a very bad signal for future US-Soviet cooperation. NSF believes it will also cast doubt on US reliability as a scientific leader and participant in global geoscience projects such as the International Geosphere/Biosphere Program. An agreement on cooperation in space may also be threatened by the muddle over ODP.

NSF is pushing for a decision from the Reagan administration before July, when a large ODP planning meeting will be held in Strasbourg. It is now up to the National Security Council, and ultimately President Reagan, to decide whether national security issues outweigh the damage to international relations that would be caused if the United States reneges on its initiation to the Soviet Union to join ODP.

Joseph Palca

West German AI venture launched

Munich

A NEW research centre for artificial intelligence will be set up at the universities of Kaiserslautern and Saarbrücken, it was announced by the West German Minister for Research and Technology, Dr Heinz Riesenhuber, on 27 March. The centre, to be administered by a group of computer companies, will start up in early 1989.

Several companies, including Siemens, Nixdorf and ADV/ORGA, have expressed an interest. The Fraunhofer Society for the Advancement of Applied Research and the Mathematics and Data Processing Research Corporation, both government research institutions, are also expected to take part. The Ministry for Research and Technology (BMFT) will support the projects to be carried out at the centre, which itself will be paid for by the companies and universities.

BMFT has paid up to 50 per cent of the costs of cooperative projects in information processing among industrial companies or between industry and academic institutions over the past three years. The budget for such projects is DM100 million per year.

A decisive element in the companies' choice of site for the new institute was the fact that about 40 people at the two universities are working on subjects related to artificial intelligence, such as natural language processing. Topics to be added include 'expert systems'.

Despite the large public support, the goals of the centre are modest. It is hoped that as many as 50 researchers will join, but, according to a source at the BMFT, "it is not a question of how many we want", but of how many qualified people can be found.

Steven Dickman

New batteries needed for design of US space telescope?

San Francisco

PROJECT managers for the Hubble Space Telescope are faced with the tough decision of whether to try to design, test and install an entirely new set of batteries before the planned launch late next year. The tight budget of NASA (the National Aeronautics and Space Administration) is one problem, the shortage of time another—but the launch date may slip.

Bertram R. Bulkin, director of scientific space programmes for Lockheed Missiles and Space Co. in Sunnyvale, California, where the 13-metre-long instrument is stored, estimates that the fix will cost about \$12 million. But the money might be well spent because the present batteries may not be up to the job.

Ever since engineers ran the telescope through a simulated "thermal vacuum" test last summer, they have fretted over the discovery that the telescope uses more power than expected, chiefly by the heaters which keep the five scientific in-

struments at operating temperature. They estimate that the telescope would be idle up to 20 per cent of the time while its solar panels recharge the batteries.

"The bird is just plain colder than we thought", said Fred Wojtalik, project deputy manager at NASA's Marshall Space Flight Center in Huntsville, Alabama. The telescope and its 2.4-m mirror was to have gone into a 600-km orbit last October with six nickel-cadmium batteries. Each has a capacity of 60 ampere-hours, and must go through 16 charge-discharge cycles a day. NiCd batteries lose 40 per cent or more of their capacity under these conditions after two or three years, making periodic replacement necessary. The new heating problem means even earlier maintenance by the space shuttle.

The project leaders are reluctantly willing to fly with the present batteries, but have embarked on a crash study of nickel hydrogen batteries—pressurized devices

operating much like fuel cells—which are standard equipment for many satellites in geosynchronous orbit, but which have never had to survive the rapid charge-discharge cycles of low-orbiting satellites passing through the Earth's shadow many times a day. They offer high output and a remarkable resistance to degradation. "They seem to get better with time", Bulkin says, and they should "last a good five years with no loss in capacity".

Nickel-hydrogen batteries could accommodate twice the peak power demand of the present batteries, allowing more instruments to run, or more frequent shifts in attitude by the observatory. The longer total life of the batteries could also obviate the need for one or two space shuttle maintenance missions.

If there are funds and if tests show that the new batteries would function properly, "then we are going to go for them", says Wojtalik. At the very least, the engineers plan to modify the NiCd battery-pack mounting points so that, even after the telescope is in orbit, astronauts could remove the old batteries and replace them with nickel hydrogen units. **Charles Petit**

US university presidents fight shy of pork-barrel politics

Washington

FACED with ageing facilities and decreasing federal support for their replacement, should research universities make individual requests directly to Congress for new buildings and equipment? In the past few budget cycles, several institutions have sought so to bypass the federal agencies ostensibly qualified to make merit judgements on such requests. The issue dominated the annual meeting last week of the American Association of Universities (AAU), and will be hotly debated this year, especially when Congress starts doling out money for 1988.

A new report, written by a committee chaired by Donald Langenberg, of the University of Illinois at Chicago, concludes that if Congress begins routinely to give money directly to institutions, it could "cause fundamental changes and damage to the research enterprise". The report argues that Congress does not fully appreciate the need for peer or merit review, and that some of its members regard the system as a merely an "old-boy" network to preserve the scientific *status quo*.

Funds earmarked for a particular university are often part of the pork-barrel politics by which individual members of Congress win rewards for their constituents. Representatives can win support for facilities which, although small on the scale of the federal government's support for science, may be a welcome windfall for

the recipient institution. The tempting possibility of easy research money has prompted many universities to hire Washington lobbyists.

The Langenberg report recommends that Congress and universities should find a compromise between political realities and strict adherence to merit review. One partial solution would be a competitive research facilities programme, as already proposed in Congress by Representative Robert Roe (Democrat, New Jersey), chairman of the House of Representatives Science, Space and Technology Committee. AAU supports the Roe bill.

But some believe, with Cornell University president Frank Rhodes, that any divergence from peer-review will lead to chaos. In December 1985, Rhodes turned down a \$10-million Defense Department grant for a new computer because it had not been subjected to peer-review. Cornell never received the money, and Rhodes must now worry that he may have damaged Cornell's chances of receiving Defense Department money in future.

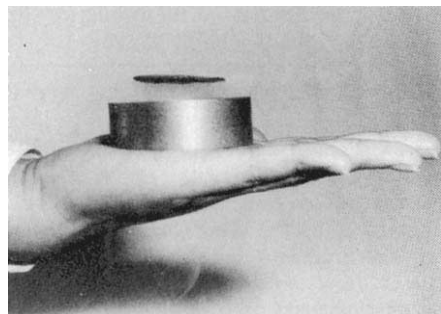
The issue of earmarking is so divisive that even the Langenberg committee members could not reach a consensus. In a dissenting addendum to the report, Arthur Sussman of the University of Chicago argues that willingness to accept even a modified form of earmarking would be "detrimental to the long-term interests of higher education". **Joseph Palca**

Superconductor rising high in Japan

Tokyo

A conjuring trick?—no, another example of the breathtaking speed at which the Japanese are developing high-temperature superconducting ceramics.

This demonstration of levitation by Mitsubishi Electric Corporation was performed by releasing a disk of barium-yttrium-copper oxide 4 cm in diameter (8 g weight) over a permanent magnet immediately after dunking the ceramic disk



A floating ring of high-temperature superconducting ceramic.

in liquid nitrogen. A current induced in the disk by the Meissner effect set up an opposing magnetic field on which the disk 'floated' above the magnet. Mitsubishi Electric claims a current of about 40 A was induced in the disk—equivalent to a current density of about 200 A cm^{-2} , nearly two orders of magnitude larger than that reported last week by Toshiba Corporation of Japan for high-temperature superconducting ceramic wire (see *Nature* 326, 533; 1987). **David Swinbanks**

Hesse election brings nuclear relief by a narrow margin

Munich

A NARROW victory by West Germany's ruling conservative coalition in the 5 April state elections in Hesse has removed one of the major obstacles to the coalition's nuclear policy.

The victory for the coalition — composed of the Christian Democrats (CDU) and the Free Democrats (FDP) — assures a conservative majority for the foreseeable future in the *Bundesrat*, the upper house of the federal parliament. Composed of representatives of the state (*Land*) governments, the *Bundesrat* has the power of veto over any laws, such as those governing the nuclear industry, that affect the interest of the states. By winning most of the five state elections being held this year in West Germany, the Social Democrats (SPD) would have threatened to tie the hands of the Kohl government.

The Hesse election, in which the CDU and FDP won a total of 56 seats in the 110-seat state parliament, ended an uninterrupted 40-year span of SPD rule in heavily industrialized Hesse. The latest SPD government, a so-called "red-green" coalition with the Green party, collapsed in February over a dispute involving the processing of plutonium at the nuclear fuel manufacturer Alkem. In March, the government of Hesse filed suit in the German Supreme Court at Karlsruhe challenging the federal government's jurisdiction over the plutonium industry and calling for a halt to the handling of plutonium at Alkem. The new Minister-President of Hesse, Walter Wallmann (CDU), has announced that he will withdraw the suit as soon as he takes office.

Wallmann's return to Hesse (he served as mayor of Frankfurt from 1977 until 1986) was greeted with a sigh of relief from industry, which had feared that a red-green coalition with a stronger Green representation would mean stricter and costlier environmental controls.

Perhaps most significant for the West German political scene was the collapse of the Social Democrats. They lost voters both to the CDU and to the Greens, whose share of the vote grew to a solid 9.4 per cent. This loss by the SPD in a state it has long controlled has been attributed to its attempt to please Green-leaning voters with a strong pro-environment policy without doing enough to retain its traditional base of unionized workers. The loss has also been attributed to the recent resignation of long-time party leader, former Chancellor Willy Brandt.

The Greens, the environmental party whose inclusion in the red-green coalition in 1985 marked a turning point in German politics, now find themselves once again in

the opposition. A demoralized Karin Guder, head of the Greens' Hessian party office in Frankfurt, said that the party's "only chance" now of influencing policy is through public information and citizen initiatives. The 3.5 per cent growth in support for the Greens compared with the last election is "small consolation" for the loss of parliamentary responsibility.

But the election results are less sourly regarded by "fundamentalist" Greens, who had disapproved of entering into the coalition in the first place.



Wallmann — "brutal politics" win through.

Jutta Dittfurth, leader of the so-called "eco-socialist" wing of the party in Frankfurt, said that it is preferable to carry out "radical reform politics" while outside the ruling coalition, especially when there is an SPD minority government, as was the case before the red-green coalition was formed. Despite their strong showing in the elections, the Greens will be hard-pressed to counter what Dittfurth called the "brutal politics" of Wallmann.

In becoming Hessian Minister-President, Wallmann vacates the post of federal Environment Minister. His replacement will be Klaus Töpfer (CDU), who had been Environment Minister of the state of Rhineland-Palatinate since 1985. Töpfer is expected to follow the nuclear policies of his predecessor.

According to a source in the Rhineland-Palatinate Environment Ministry, Töpfer's approval of a gradual "renunciation" of nuclear energy merely reflects the CDU nuclear programme, which calls for the continued reliance on nuclear energy until alternative energy sources have been found. The conservative *Frankfurter Allgemeine Zeitung* newspaper called Töpfer a "master of damage control".

Steven Dickman

New research centre

London

THE General Electric Company (GEC) is to build a new research centre at the Cambridge science park. An eight-acre site has been acquired from Trinity College, the science park trustees.

The unit will be completed in three years' time, says Dr John Williams, managing director of GEC Research, and will create 600 jobs (500 for graduates).

In moving to Cambridge, GEC acknowledges the success of the science park as a centre of excellence and the benefit of strengthening their links with the university. This development is part of the fourth of five phases of growth at the park.

The centre will have a multidisciplinary approach with research initiatives in areas such as microelectronics as well as information technology and software, and is expected to provide technological advances for the company's existing production units.

S.H.

Canadian Ariane launch

London

TELESAT's next communications satellites, Anik E1 and Anik E2, will be launched in 1990, and will be the first communications satellites launched by Arianespace. A launch agreement was signed by Eldon Thompson, President of Telesat Canada, and Frederic D'Allest, chairman of Arianespace, on 10 April. Ariane is one of the most powerful European launch vehicles and places satellites directly into geostationary orbits. "The dedicated launch will ensure that each satellite will have an operational life of ten years" said Thompson. Telesat now has five Anik satellites in orbit, more than any other domestic satellite company. The new satellites are designed to meet their communications requirements until at least the year 2000.

S.H.

Rees to be MRC secretary

London

DR David Allan Rees will be the first non-medical to be secretary of the Medical Research Council (MRC) when he succeeds Sir James Gowans on 1 October this year.

Rees, whose research interests have been in the structure and biochemistry of polysaccharides, is at present director of the MRC's National Institute for Medical Research and visiting professor in biophysics at King's College, London. He has instigated a major reorganization at the institute and created an MRC Collaborative Centre where council and industrial scientists can work side by side. His successor has yet to be selected.

It is understood that Gowans will continue to organize the MRC-financed AIDS (acquired immune deficiency syndrome) programme at least until the end of the year.

S.H.

Soviets woo cream of Indian science by multi-faceted projects

New Delhi

THE Soviet Union is wooing Indian scientists in a big way, and Indians, who traditionally have looked to the West for inspiration, will soon build connections with the Soviet Union in both basic research and high technology. The change was signalled last week with the signing of a document for extensive and long-range cooperation in science and technology.

Great importance is attached to the new agreement because of the political backing it has been given by the Prime Minister, Rajiv Gandhi, and the Soviet leader Mikhail Gorbachev. A major science initiative was at the top of the agenda of talks between the two leaders when Gorbachev visited New Delhi last November. The document signed on 21 March is the follow-up to these talks and is expected to mark the beginning of a new era of more science cooperation.

The document, described as a working paper, was the outcome of week-long discussions between a high-level Soviet delegation and leading Indian scientists led by Professor C. N. R. Rao, chairman of the prime minister's scientific advisory committee. The Soviet team, headed by Guri Marchuk, president of the Soviet Academy of Sciences, included 22 academicians, a Nobel laureate and 16 experts — the largest and most powerful group of scientists to visit India. This has been interpreted as a proof of the Soviet wish to put science cooperation, now limited to sporadic exchanges of scientists and joint work in a few areas such as energy and meteorology, on a higher plane.

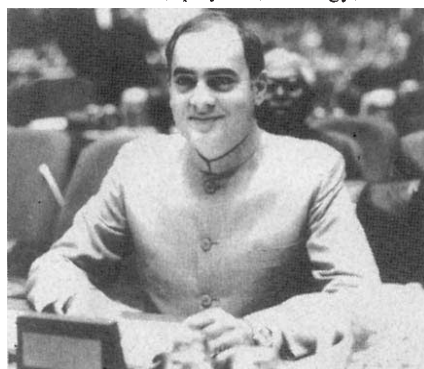
The bear-hug is also regarded by some scientists as a move by the Soviet Union to reduce Western influence on Indian science. "A lot of Indian scientists go to the United States to work and contribute to its economy, and the Soviet Union would like to divert this flow", said one Indian physicist who took part in the negotiations.

In their enthusiasm to impress the Indians, the Soviets initially proposed 200 projects in 17 areas for collaboration. India considered the offer unmanageable and reduced it to six areas where results could be seen in 3–5 years. Specific details will be worked out during the reciprocal visit of an Indian delegation to Moscow in June.

The Soviet Union will help to set up a commercial facility for production of medical and industrial lasers. In materials science, collaboration is planned on powder metallurgy, processing of alloys in space and synthetic diamonds. The Soviet Union will launch two more Indian remote-sensing satellites and will provide

technology for the development of a space-borne microwave sensor. A unique technology developed by the Soviet Union for locating ground water using nuclear magnetic resonance will be shared with India.

In biotechnology, joint research is planned on AIDS (acquired immune deficiency syndrome) and diagnostic kits for communicable diseases. The Soviet Union will provide the technology for building a proton synchrotron in India. The two countries have also agreed to strengthen cooperation in basic research in mathematics, physics, biology, chem-



Gandhi considers Soviet science initiative.

istry and oceanography. Cooperation in computer electronics may be included later. India had Soviet help in the development of an orbiting aerospace vehicle, and the Soviet Union has offered to set up an international space centre in India. According to Rao, these ventures will be discussed by the Indian space department separately with the Soviet Union.

The proposed collaboration will involve 60 research institutes in India and in the Soviet Union and an exchange of 200–300 scientists every year. Indian nuclear physicists have already visited the synchrotron laboratory at Novosibirsk, and another team is due to visit Turkistan to learn about Soviet technology for constructing dams over seismic faults.

What has impressed Indian scientists is the Soviet readiness to open up its research institutions on an unprecedented scale. Indians have been invited to take their families and work for extended periods in the Soviet Union. Marchuk said a large number of Indian students would also be welcome for graduate and post-graduate studies in Soviet universities.

Whatever the motive of the Soviet Union, Indian science policy-makers feel that a close encounter with the Soviets would be to the advantage of Indian science. "There are several areas in which the Soviet Union is very advanced", said Rao, "and we are very enthusiastic about working with them." **K. S. Jayaraman**

Soviet reforms stop careers in mid-stream

London

THE rejuvenation of the Soviet Academy of Sciences appears to have brought several careers to a premature end. In an interview with *Izvestia*, the newly appointed president of the academy, Dr Guri Marchuk, said that 416 holders of the degree of Doctor of Science, or one in ten of those on the staff with this most senior of the Soviet science qualifications, had failed a recertification test.

According to Marchuk, the academy has carried out a "reassessment of academic qualifications in the light of current needs", as a result of which 2.5 per cent of scientists at academy institutes have been dismissed and a further 5 per cent demoted. Those with the DSc degrees who had failed the test had been "creatively active" people who had once organized laboratories and research teams, but who were now living on their past reputations. Marchuk said that "timely withdrawal" of such people into purely scientific work must in future be the norm.

Under the academy's new policy, people will have to retire from administrative posts at 65, but retired directors of research institutes have the option of continuing as advisers or in research, and will be encouraged to write books to "sum up their experience". Marchuk said that there will be an annual "rejuvenation" of the academy's institutes "without increasing their numerical strength".

The fate of those thus shed from the academy's institutes is unclear, but present policies aimed at closer integration of research and industry would suggest that many will end up in the "sectoral" institutes run directly by the production ministries, which have proliferated during the past 25 years, often for no obvious reason other than to bolster the prestige of the sponsoring ministry.

Many of these institutes are inadequately staffed and equipped, and there is considerable duplication of work. Mr Mikhail Gorbachev's proposals for the restructuring of Soviet science urge the closure of the least productive of these institutes, the amalgamation of others and greater supervision by the academy of the work of the sectoral institutes as a whole.

Marchuk's own position is that the closure of the academy's technical department at the end of the Khrushchev era was "unjustifiable", but that in present circumstances the academy is not the place for industrial research. He said that many of the industrial ministries have research budgets greater than the academy's entire budget. **Vera Rich**

Congress gives supercollider easy round but worries about cost

Washington

The proposal to build the \$4.4 billion Superconducting Supercollider (SSC) is facing its first major test since President Reagan gave it his blessing in January. After three days of hearings by the House Committee on Science, Space and Technology beginning on 7 April, it was clear that SSC is respected for its scientific merits but that there are worries about the cost.

And, as a sign of things to come, testimony from various states' spokesmen, ostensibly to demonstrate the breadth of general support for the SSC, quickly turned into an unashamed procession of pitches for particular sites.

The first day of the hearings was mostly devoted to science, with an array of witnesses including Steven Weinberg, Leon Lederman (director of Fermilab), Herwig Schopper (director of CERN), Robert Schrieffer (of the BCS theory of superconductivity) and more. But science was not uppermost in the Committee's mind.

Alvin Trivelpiece, outgoing director of Energy Research at the Department of Energy (DOE), was questioned for almost three hours on the cost of SSC, its effect on other research, the significance of the new superconductors and (clearly most important for some congressmen) the logic of the site selection procedure.

In their opening statements, many committee members alluded to the general benefits of large science projects. The falling quality of science education and the faltering competitiveness of high technology industries were frequently cited as reasons for giving US science a helping hand. But Dan Ritter of Pennsylvania thought the money would be better spent directly on applied research, especially now that high temperature superconductors are causing so much excitement. Reports that the Japanese government will try to build a liquid nitrogen cooled superconducting magnet in two years seemed to him a reason to delay the SSC for a while.

Maury Tigner, director of the SSC Central Design Group, did not believe that high-quality magnets could be made using the new superconductors in less than ten years, and therefore saw not a reason to delay but rather a hope that the SSC might be upgraded to substantially higher energies, perhaps in twenty years. In any case, removing the liquid helium cooling systems would save only about 5 per cent of the construction and running costs over the first decade of operation.

Many of the scientific witnesses agreed with Trivelpiece that the SSC should not go ahead at the expense of 'small science'.

But only James Krumhansl, professor of physics at Cornell University, was openly hostile to the project. After saying that he found the SSC "truly exciting", he argued that cutbacks in other areas were an inevitable side effect. Because annual reviews of science expenditure are mandatory, there is no way that funds can be set aside for one project over a period of years.

The most insistent questions during Trivelpiece's testimony concerned site selection. The official invitation for site proposals was issued on 1 April, with a deadline for submissions of 3 August. This was felt to be too short a preparation period by many congressmen, especially those from states which had followed DOE's advice and not begun a site search in earnest until after presidential approval had been given.

Committee member Tim Valentine of North Carolina complained that the pro-

cedure did not allow his state and many others to match the already well-researched efforts of big spenders such as Illinois, Texas and California, although his objection was somewhat mitigated the next day when Governor James Martin of North Carolina presented testimony which went from praise of the SSC to a summary site proposal, mentioning the different kinds of rocks in his state, the nature of groundwater distribution as well as the pleasant way of life to be had there.

Spokesmen for Illinois and California, on the other hand, said that anyone with a serious interest in the SSC would have known two years ago that a site invitation was on its way. Illinois has already spent \$4.5 million on site research, and is proposing to spend \$15 million more to advance its claim that the SSC should be built near Fermilab, which it would use as an injector.

Trivelpiece and all the SSC supporters hope that site selection can be carried out agreeably. Too much bickering between the states could result in widespread political support for the project collapsing.

David Lindley

Japan takes action to control trade of endangered wildlife

Tokyo

JAPAN is at last taking legislative action to control the illegal import of endangered wildlife. But conservationists say the new bill, expected to be submitted to the current session of the Diet, will fail to halt most of Japan's illegal trade.

Japan signed the Washington Convention on International Trade in Endangered Species of Wild Fauna and Flora in

however, to stipulate the return of illegally imported wildlife to the country of origin.

Tom Milliken, director of TRAFFIC Japan, the trade monitoring area of the World Wildlife Fund, also points out that nearly all the species covered by the new bill are Appendix 1 species, trade in which is totally banned. But by far the greater volume of illegal trade in Japan occurs in Appendix 2 species which are not covered by the new legislation.

Reservation species are also excluded from the new law. On signing the convention, Japan exempted 14 Appendix 1 species — more than any other party to the convention — including six species of whale, three species of monitor lizard, the saltmarsh crocodile and the Himalayan musk deer.

Finance Ministry records show that last year Japan imported 367 kg of musk, 90 per cent illegally smuggled from China. One kilogram of musk fetches about ¥10 million (\$70,000) on the Japanese market but requires the slaughter of more than 100 animals. The musk deer has been wiped out in most of India and is found only in a few isolated areas of the Himalayas. The musk is used to make perfumes and a traditional Chinese aphrodisiac.

Under strong pressure from abroad, MITI and the Ministry of Health and Welfare jointly agreed in February to remove the musk deer from the reservation list... by March 1989.

David Swinbanks

COMPASSIONATE JAPANESE OFFICIAL (ENDANGERED SPECIES)



1980. But widespread flouting of the convention prevails because Japan has no domestic legislation to back the convention's provisions.

The new bill bans the import of 492 species of fauna and flora and provides for a maximum fine of ¥300,000 (\$2,000) or a six-month sentence for violators. The legislation also empowers the authorities to enter pet shops and zoos to make on-the-spot investigations. The bill fails,

Reform urged for "hellish" Japanese education system

Tokyo

JAPAN'S National Council on Education Reform has handed over its third set of recommendations to Prime Minister Yasuhiro Nakasone. The council calls for a more open and flexible school system, the promotion of lifelong education, and more funds for the national universities. But there are no radical proposals to deal with the major problems besetting Japan's higher education system — the pre-university examination "hell" and elitism in the universities — and the Ministry of Education's strict control over school textbooks will be maintained.

The council was set up by Nakasone in 1984 to map out the first major reforms of Japan's education system since the post-war US occupation. Two sets of recommendations have already been issued, in June 1985 and April 1986 (see *Nature* **318**, 7; 1985; and **319**, 347; 1988). But from the outset the council has been torn by internal strife and several of the more sweeping reforms put forward by council members have been dropped or shelved for further discussion.

In its latest proposals, the council calls for increased public and private sector support for the universities (private and national), by establishing financial links with local governments, instituting tax incentives for donations by private individuals and organizations, and by encouraging national universities to establish their own foundations with such donations and to use effectively land they own for investment purposes. But the council stops short of an earlier proposal by one of the council's subcommittees to convert national universities into corporations — a decision on this will be made before the council dissolves in August.

A university tenure system is recommended, instead of life-time employment, with some staff being hired on a contract basis to increase fluidity in academic appointments. And qualifications criteria will be relaxed to allow employment of more non-academic and foreign staff.

A major thrust of the latest recommendations is the promotion of lifelong education. The council calls on universities to open their doors to adult job-holders at both the undergraduate and graduate level, and local governments are recommended to establish "intelligent" adult education facilities to be open around the clock and contain advanced computers.

But there is little to be found in the recommendations to solve the intense competition at the high school level. Senior high schools are encouraged to consider achievement reports from junior high schools, interviews and composi-

tions, and to place more emphasis on the ability to think and solve problems, instead of rote learning, in their entrance exams. But the system of after-hours cram schools will be maintained; the council merely recommends that the central and local governments supply parents with accurate information on this private education industry, which is expanding by leaps and bounds.

One enterprising cram school, Gakkyusha Co., plans to launch nationwide television lessons via a communications satellite to be launched by Japan Communications Satellite Co. next year. Others have started summer cram camps in winter ski resorts, squeezing students into every available space, including the ski-drying room, to prepare them for the highly competitive spring entrance examinations for junior and senior high schools.

Rigid control over what these school pupils read will be maintained by the Ministry of Education, Science and Culture. The council's first subcommittee on reform for education for the next century had called for abolition of the ministry's textbook screening system, which has been strongly criticized by neighbouring nations for toning down descriptions of Japan's militaristic past. But the subcommittee on primary and secondary education, which is close to the ministry, and a committee of the ruling Liberal

Democratic Party, including three former education ministers, opposed the move. The screening system will be simplified from a three-stage to one-stage process, but critics say this will merely increase the ministry's control over textbooks.

A minor change in the "hell" of university entrance examinations has already been instituted. This year candidates were allowed to sit entrance examinations for two or three national universities, instead of only one. This was done by dividing the universities into three groups according to area and holding the examinations for each group at different times.

In the past, national university entrance examinations were all held on the same day, each university having its own examination on its own campus. Japan's young university aspirants leapt at the new chance, commuting back and forth on the bullet train, textbooks in hand, in this their first opportunity to sit examinations for both of Japan's most prestigious universities, Kyoto and Tokyo.

The result was a dilemma for the universities. How many applicants should they accept to fill their quota? Kyoto's famed science faculty undershot the mark, accepting 485 applicants for 291 places. Hours before the application deadline on 25 March only 220 had accepted, the rest presumably having chosen Tokyo, and telegrams were hurriedly sent out to additional applicants to fill the quota. But although Kyoto University's pride may have been dented, there is no indication that the competition for entrance to Japan's most elite universities has subsided.

David Swinbanks

Falsified data help to keep Soviet metal pockets lined

London

SEVERAL Soviet metallurgical plants have been accused of falsifying metallurgical data on a new tungsten-free steel to safeguard their own pockets. The accusation, made during a broadcast of Moscow television's popular *Science and Life* programme last month, is that the new steel has been falsely represented as unsatisfactory to keep it off the market.

Under the traditional system of financial rewards in the Soviet Union bonuses and other incentive payments are determined by gross turnover. The search for tungsten-free steels was begun some years ago but, according to Moscow television, the new steels would have been so much cheaper than those using tungsten that the turnover of the machine-tool fabrication plants using them would have lost money, on paper at least.

Institutes of Azerbaijan Academy of Sciences appear to have played an important part in the development of the tung-

sten-free steels, partly through a special design bureau known as Kristall. N. M. Suleiman, director of the Kristall bureau, told television journalist B. V. Goldaev that there had been opposition to the new steels in the industry and that, on some occasions, the results of laboratory tests had been falsified. But, he said, tests carried out by industrial enterprises, especially the KamAZ enterprise, had shown that tools made from the new steel performed well. Even so, Suleiman said, the Soviet Ministry of the Machine Tool and Toolmaking had ignored the new steel.

M. A. Stremel, a technologist, pointed directly to the economic paradox that, whatever the merits of the new steels, "no metallurgical works is going to ruin itself" by producing them at a third of the price of conventional tungsten steel. A reform now at the draft stage would link performance to net rather than gross turnover of enterprises, but is not expected to be in force generally before 1990. Vera Rich

Texan research centre struggles to survive oil money slump

Houston, Texas

A CLUSTER of buildings in a verdant development about 35 miles north of Houston is the home of a new multidisciplinary research institution trying to bring the talents of Texan universities to bear on the needs of industry. Certainly the new centre, called HARC for Houston Area Research Center, has attracted interest. But interest is not money and in the depressed state of the Texan economy, HARC is scrambling to avoid crashing before it can take off.

HARC began in 1982. The idea was to provide "an innovative research and development environment" that government and industry could take advantage of. As well as developing its own resources, HARC works in liaison with members of the faculty at Rice University, Texas A&M University, the University of Houston University Park campus and the University of Texas at Austin.

George Mitchell, chairman and president of Mitchell Energy and Develop-

ment, has provided the inspiration — and a lot of the cash — to get HARC off the ground. Mitchell is also the driving force behind The Woodlands, the planned community in which HARC is located. In 1984, Mitchell Energy and Development Corporation, The Woodlands Corporation and Mitchell himself together provided HARC with \$2.4 million, nearly half of HARC's cash income. In 1985, Mitchell's share rose to \$2.6 million, although total income rose to nearly \$8 million, mainly because of a Department of Energy contract to the Texas Accelerator Center, one of HARC's eight divisions.

But Department of Energy support for the Texas Accelerator Center has evaporated with the decision to use superconducting magnets, rather than the superferric design developed by the Texas group, for the Superconducting Supercollider (SSC). There is still some bitterness about the magnet selection process. Three completed magnets sit in the centre's workshop, testifying to the feasibility of the design.

Shut out for the time being from SSC projects, the accelerator centre has looked in other directions for financial support. One promising project is to apply the expertise gained in designing the SSC magnets in the design of magnets for magnetic resonance imaging devices (MRI), formerly called nuclear magnetic resonance imagers. HARC has received funds from Baylor College of Medicine to build a magnet for an MRI, and there may be a large commercial market for a new version of the increasingly popular MRIs.

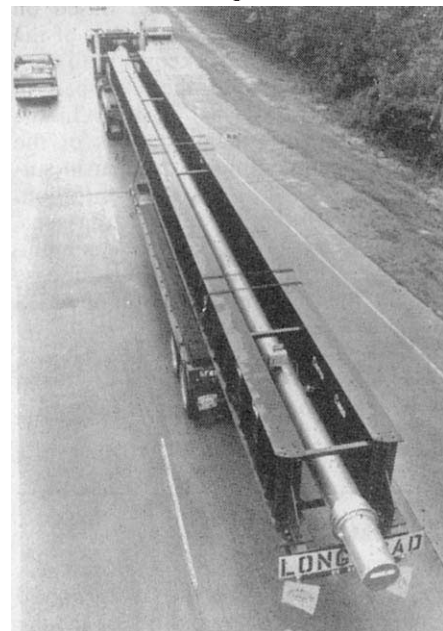
HARC also hopes that its new SX-2 supercomputer, manufactured by Nippon Electronics Corporation (NEC), will enable it to compete for contracts. HARC took delivery of the new computer last summer, and is still working the bugs out of the system.

Even so, HARC has been criticized for deciding to purchase a \$20-million Japanese supercomputer. HARC's purchase appears to have sparked the trade investigation on the trade in supercomputers launched last year by the US Trade Representative. HARC vice-president for development, Jane Armstrong, says HARC knew it would run into trouble for the decision. But, she says, HARC based its decision on strong commitment from NEC to support it. NEC has already awarded HARC a \$500,000 contract, and there is promise of more to come.

HARC also believes there is great potential for the future in its Space Technology and Research (STAR) Center; it announced last month that STAR had formed a partnership with Space Industries, Inc. to provide equipment and

experiments for the Industrial Space Facility. David Norton, STAR's director, says this agreement will give HARC access to space before the space station is operational. Norton is enthusiastic about the commercial possibilities that a micro-gravity environment may provide. But no cash will change hands in the deal with Space Industries: HARC will instead receive space on the new facility in exchange for its development efforts.

Perhaps the most promising area for commercial development at HARC comes from the Geotechnology Research Institute and its energetic director Manik



Superferric magnet on its way to storage.

Talwani. Talwani had been director of Columbia University's Lamont-Doherty Geological Observatory and after that director of research for Gulf Oil. He would like to establish an energy research consortium, similar to the electronics industry's Microelectronics and Computer Technology Corporation (MCC). Such a consortium would focus on recovery techniques to bring down the price of recovering oil from fields that have already yielded up their easily retrievable reserves.

Talwani argues that the depressed oil economy may paradoxically help to generate interest in HARC. Oil companies that have had to scale down their own research and development activities may find it cost-effective to contract for such services with HARC. Talwani is also trying to put together a \$22 million programme for deep reflection profiling on the continental margins. The National Science Foundation is said to be interested.

The depressed oil economy has made Texas industry less willing or able to support new ventures such as HARC. If it can scrape by through the hard times, HARC may nevertheless provide a model for future relations among industry, universities and government.

Joseph Palca

Locust plague ahead?

London

ANXIETY about a new locust plague in the wake of the impending rainy season in Africa has prompted attempts to resuscitate international programmes for locust control, many of which have been run down or even abandoned in the several years since the last serious plague of locusts in Africa. With the breaking of the African drought in 1985, conditions are now thought to be more favourable for a plague of locusts than at any time for 50 years. High-density populations of the four species of locusts in Africa are known already to be attacking vegetation.

Preparations to head off what could be a plague affecting the whole of Africa are being coordinated by the Emergency Centre for Locust Operations of the UN Food and Agriculture Organization, which last year raised \$51 million by an appeal for funds for locust control. This year, \$19 million has been pledged so far, while the British government has promised an extra £7 million.

Last year's programme of ground and aerial control seems to have headed off what could have been a plague. Small-scale control programmes have already begun, with farmers applying pesticides to protect their ground from newly hatched insects. After the expected rains, control will focus on aerial spraying of large areas. Roger Bone of the Tropical Development and Control Institute says the situation in Africa is "still dangerous". Kathy Johnstone

AIDS caused by a slow virus

SIR—Recognition that AIDS (acquired immune deficiency syndrome) is caused by a lentivirus has been slow in coming¹. Ellrodt and LeBras² do a signal service to those devising vaccines for human immune deficiency viruses (HIVs) by offering a cautionary warning about the unusual properties of these slow viral agents. Similar warnings should be issued to policy planners in their projections of future prevalence of AIDS. Based on other lentiviruses, the full impact of the HIVs may not become known until one-half to two-thirds of the life span of infected individuals has passed³. In the case of humans, manifestations of the disease could continue to appear in survivors for 20–40 years after infection. Coupled with the prospect of transgenic humans (T.M. Folks, personal communications) arising over this long disease course, the case for prevention of AIDS seems even more pressing.

CECIL H. FOX

National Institutes of Health,
Bethesda, Maryland 20892, USA

1. Connor, S. *New Scientist*, 12 February (1987).
2. Ellrodt, A. & Le Bras, P. *Nature* 325, 765 (1987).
3. Haase, A.T. *Nature* 322, 130–136 (1986).

Soviet AIDS

SIR—There has been official frankness about the presence of AIDS (acquired immune deficiency syndrome) in the Soviet Union for rather longer than Vera Rich gives credit for (*Nature* 326, 3; 1987). When the agreement between Britain and the Soviet Union on cooperation in the field of medicine and public health was resigned last November after an almost total lapse of seven years, a protocol was drawn up providing for cooperative research in selected fields, including public health and laboratory aspects of AIDS. Specifically covered were viral cultures and characterization of strains, including molecular structure, and studies of methods of detection of antibody, distribution of antibodies and means of epidemic control. There was no suggestion that this research was of a purely academic interest to the Soviet Union.

The castigation by a Soviet deputy minister of health of Western companies that keep secret their work on AIDS diagnostics may at first sight appear to be mere propaganda against the greedy, capitalist West. But to anyone raised on a diet of Communist teaching, some Western commercial practices must seem highly unethical. We have been schooled in the merits of free enterprise and can rationalize that to give a competitor a costless lead would mean less profit for, and ultimately less research by, the pioneering company.

Yet even then, not all of us are immune to doubts as to the ethics of full commercial secrecy where potentially life-saving products are being researched and developed and when one of its results is colossal duplication of scientific effort.

STEWART BRITTEN

(Hon. Secretary, UK-USSR Medical
Exchange Programme)

9d Stanhope Road,
London N6 5NE, UK

Rabies experiment

SIR—We wish to correct a number of misconceptions and inaccuracies in connection with the rabies experiment by the Pan American Health Organization (PAHO) in Argentina, which was the subject of a letter from an independent group of Argentinian scientists (*Nature* 324, 610; 1986).

Wistar's part in the experiment was solely as consultant for the development of the protocol, fund-raiser for the costs of the experiment and supplier of the vaccine to PAHO. But as you refer to the Wistar Institute in the title of the letter, we wish to clarify the actual conditions of the experiment.

The experiment was designed to evaluate a live recombinant virus vaccine for immunization of cattle against rabies. The vaccine, a vaccinia-rabies glycoprotein (V-RG) recombinant virus, had previously been tested in animal species in captivity, including mice, rabbits, foxes, skunks and raccoons. In each case, protective immunity upon challenge with rabies virus had been demonstrated.

Before the experiment in Argentina, the level of rabies virus-neutralizing antibodies in cattle inoculated with the V-RG vaccine by intradermal, subcutaneous and intramuscular routes had been shown to be compatible with protection of these animals against a lethal challenge dose of rabies virus (P. Desmettre, personal communication). As the Argentine scientists raise the question of "possible alteration in the tissue localization of the recombinant virus", it should be emphasized that in all previous experiments with other animal models in which the spread of V-RG recombinant virus within the body was monitored, there was no evidence of gross or histopathological lesions caused by virus infection in any tissues collected *post mortem* and no evidence of increased neurotropism was detected in laboratory animals inoculated with varying doses of V-RG recombinant virus by a variety of routes. This would indicate that any vaccinia virus-rabies glycoprotein pseudotype that might emerge does not enhance the already reduced neurovirulence of V-RG recombinant virus.

In the vaccine trial in Argentina, 40 cattle were divided into two groups. On 11 July 1986, ten animals in one group were

vaccinated subcutaneously in the neck with 1 ml of the V-RG recombinant vaccine administered from a 'multiple dose' syringe, similar to ones used for routine vaccination of cattle. In the other group, ten animals were vaccinated with the recombinant vaccine by intradermal scarification in a previously shaved area on the neck of each animal. The ten cattle vaccinated subcutaneously and the ten cattle vaccinated by scarification were kept in separate enclosures but in close contact with ten non-vaccinated cows. The enclosures for each group of 20 animals (ten vaccinated and ten non-vaccinated) were separated by a distance of a quarter of a mile and consisted of a shed in a fenced field. During the first month of the trial, the animals were kept in the shed to ensure the closest possible contact between vaccinated and non-vaccinated cattle. Milking cattle were used in the experiment so that milk production levels could be monitored as a sensitive indicator of animal health. None of the vaccinated cattle showed any decrease in milk production throughout the trial period and none showed any sign of illness.

By order of the animal disease authority in Argentina, all 40 cattle were killed on 14 November 1986, and the blood samples obtained 3–4 months after vaccination have so far been barred from use in assays for antibodies against rabies and vaccinia viruses. The one-month post-vaccination samples had, however, already been tested, and the results indicated that the vaccinated cattle developed protective levels of anti-rabies antibody. None of the non-vaccinated animals developed anti-rabies antibodies, which suggests that the vaccinia-recombinant virus did not spread from vaccinated animals to non-vaccinated animals.

From a scientific point of view, it makes no sense for the animals to have been killed four months after immunization, when they were healthy and unlikely to have been shedding the virus, thus preventing the acquisition of further valuable information about the efficacy of the recombinant vaccine in cattle. In their letter to *Nature*, the Argentinian scientists stated: "We are not against the development of recombinant vaccine technology in view of its potential value for effective disease control". If this was said in all sincerity, what is their view of the elimination of 40 healthy animals by order of the authorities and the consequent failure to complete an important scientific experiment. Is that science or politics?

HILARY KOPROWSKI, ESTEBAN CELIS,
PETER CURTIS, BERNHARD DIETZSCHOLD,
CHARLES RUPPRECHT, MARIA TOLLIS,
WILLIAM WUNNER
Rabies Unit, Wistar Institute,
36th Street at Spruce,
Philadelphia, Pennsylvania 19104, USA

Half-truths make sense (almost)

The latest eco-policy declaration on the greenhouse effect (by the World Resources Institute) is at once (from an excess of innocence) too scary and (from ignorance) too soft.

THAT there must be a greenhouse effect is now beyond dispute; put more carbon dioxide into the atmosphere and the temperature on the surface of the Earth will inexorably increase. This simple truth is now so widely known that it is taught, along with the legends of birds and bees, in elementary schools. Television companies often broadcast the news when they are able to find a new angle on what is, by any standards, an absorbing issue in geopolitics. The fact that real people on the surface of the real Earth outside the Tropics still shiver in their winters (and even, sometimes, in their summers) seems neither here nor there.

Until now. Last week, the World Resources Institute, which describes itself as a "policy research center", and which is conveniently sited in Washington, DC, put out a stirring document, *A matter of degrees*, calculated to shiver the timbers of those who read its prediction that, in the next forty years, the average temperature on the surface of the Earth will increase between 1.6° and 4.5°. For the convenience of readers with more immediate issues on their minds, the institute has isolated these key declarations in large italic type.

Among eco-policy research centres, the World Resources Institute is one of the most respectable. Its board of directors includes not merely Mr Robert McNamara, once US Secretary of Defense and president of the World Bank, and Mr Russell Train, once chairman of the US Council on Environmental Quality, but Dr Martin Holdgate, a British civil servant who is chief scientist at the British government's Department of the Environment. Moreover, the institute is judiciously circumspect in its declaration of objectivity — how to "meet basic human needs and nurture economic growth without undermining the natural resources and environmental integrity on which life, economic vitality and international security depend?" No dropping out there.

So how can a project to warn the world of a serious danger have gone so seriously awry? The committee meeting at which the seeds of intellectual disaster were sown is easily imagined. *The greenhouse effect threatens our environmental integrity?* Right. *We must warn the world.* Right on! *But we mustn't cloud the clarity of the message with all that uncertainty?* Right again. *So what should we tell people they must do?* Don't be prescriptive; tell them

about the problem and then the market (or, in the last resort, the US Congress) will find a remedy.

As many readers of the document and of *Nature* will know, the solution of the greenhouse problem may not be that simple. The carbon cycle on the surface of the Earth predates life, not to mention human life. Even in the Precambrian, atmospheric carbon (methane, for example) was being converted into insoluble CaCO₃; photolysis made hydrogen (which escaped) and oxygen, with its propensity for chemical reaction. Since the arrival of plants, the naturally reducing environment of the Earth's atmosphere has been, against expectation, oxygenated; the flux of oxygen into the atmosphere as a by-product of the darwinian role of chlorophyll has exceeded the rate of solution in the oceans. But for the past 600 million years (and the whole duration of mammalian evolution), this state of affairs has been a transitory phenomenon. People would be well within their rights to be asking when (and how) it would end. The simple answer is when the surface of the Earth is kinematically dehydrated.

The recent anthropogenic intervention in the carbon cycle is puzzling for different reasons. For more than a century, the annual addition of CO₂ to the atmosphere has been uncomfortably greater than the natural rate of production by photolytic oxygen. The atmospheric concentration of CO₂ has increased by close on 10 per cent (to more than 320 million p.p.m) in about a century. One difficulty is that the extra CO₂ is only about half of that corresponding to the extra fossil fuel burned during that time. What has happened to the rest? Another is that the measured increase of temperature on the surface of the Earth is not as great as would be expected even from the amount of CO₂ accumulated in the atmosphere, although it is beginning to be measurable. Then there is the question of how climatic models can claim credibility while failing to handle real clouds, rather than average cloudiness.

These are among the trials with which policy research centres must contend. Tell the people that there is a muddle, or give them a clear message that they must man the barricades? Ingeniously, the World Resources Institute has found a cop-out in between these unpalatable alternatives, which passes under the name of the *model of warming commitment*: inspection

shows that "commitment" implies no more than that, if present practices are modified a little and then projected into the future, disaster (global warming) will strike *y* rather than *x* years ahead. The World Resources Institute talks of the year 2030 as critical. Its modelling takes account of other greenhouse gases than CO₂, methane for example.

Sadly, this superficial attack on the problem of global warming will carry less weight than it might have done. If those concerned had been willing to explain the uncertainties in their calculations, their gloomy predictions would have been more telling. As it is, it will be left to the rest of us to explain why it is that, after the threat of AIDS (on a timescale of a century or so) that of the greenhouse effect is one of the next in line. How might that group of problems be tackled?

There are several obvious impediments to the pursuit of rationality. The most obvious is that the advantage that accrues to separate nation states from burning fossil fuel is usually an increase of Gross National Product, but that the disadvantage (global warming) is both delayed and unattributable. It is a problem of what the environmentalists call the "global commons". Empiricists claim that short-term measures, such as the Vienna Convention on the Protection of the Ozone Layer, are valuable chiefly because they point the way to the more difficult international regime that lies ahead.

Luckily, people have been brooding on these questions for some years. At least it is possible to guess where the difficulties lie. As the World Resources Institute says, CO₂ production is a function of economic activity, from which it follows that a page-1 solution of the problem is a system of quotas to regulate CO₂ production. Predictably, there are difficulties about the partition of the allowable amounts among different kinds of producers: do developing countries have their quotas tied to what they would be producing if they were rich? Or is it that quotas will be determined by present practice?

In the longer term, say on a half-century timescale, it will be necessary to modify present practices for the generation of electrical energy. Is the only alternative nuclear fission? And, if so, what will the dissenters give up instead? These are the questions rendered inescapable by the greenhouse problem, which must sooner or later be tackled.

John Maddox

Supernova examined by computer model

THE uncertainty discussed here last week as to whether Sanduleak (Sk)-69 202 was the progenitor of supernova SN 1987A continues unresolved. Dave Arnett of the University of Chicago will no doubt be watching developments closely. He has been using computer models of stellar evolution to find out what sort of explosion would best explain the various observations — in particular its faintness and unusual light curve — and by thus working backwards his theoretical candidate for the supernova progenitor is a blue supergiant within about half a magnitude of Sk-69 202.

Arnett's success removes several difficulties. A standard type II supernova explosion was excluded for at least three reasons. Its progenitor, a red supergiant, would have been visible on optical plates of the region; the outburst would have been several magnitudes brighter than SN 1987A; and the time interval between the neutrino pulse and the optical sighting would have been longer than was the case.

These problems indicated that a smaller, fainter star would make a better progenitor. (Neutrinos emerge promptly from the collapsed core, but photons must fight their way out through the dense surrounding material; a smaller progenitor has, for this simple reason, a faster optical rise time.) But a smaller star would generate insufficient ^{56}Ni , whose radioactive decay is thought to power the recently observed brightening of the expanding shell. So some other source, perhaps a pulsar, would have to be added. Things were becoming complicated, and unconvincing.

Arnett's model is pleasingly straightforward. He takes a 15-solar mass star, gives it element abundances appropriate to the Large Magellanic Cloud (LMC), and lets it evolve. Lower elemental abundances compared with our Galaxy make the crucial difference; the central nuclear reactions are less efficient, so the star has to be hotter and denser, and therefore smaller, to maintain radiation pressure against gravitational collapse. Equally, the star is more prone to core collapse, and explodes as a less highly evolved blue star, not as an old red giant. And as a bonus, this massive star synthesizes enough ^{56}Ni to explain the later part of the light curve.

Happily, Arnett can also explain why supernovae are rare in irregular dwarf galaxies such as the LMC. Low element abundances are characteristic of these galaxies, so all 'type II' supernovae are fainter than usual, and we see them less easily.

Although there is still uncertainty over the identification of the progenitor, things are looking clearer — at least for this week.

David Lindley

Superconducting ceramics

On the way to a new technology

Noboru Miura

SINCE the recent discovery of a high-temperature superconductor with the transition temperature, T_c , exceeding 30 K by Bednorz and Müller of IBM, Zurich, and the subsequent experiments by S. Tanaka's group in Tokyo last year, tremendous efforts have been made by scientists throughout the world, in particular to achieve even higher transition temperatures. T_c now exceeds 90 K, well above the boiling point of nitrogen. The new materials are oxide ceramics, a completely new class of superconductors. Superconductors carry electric currents with no energy loss, so that they are ideal materials for making high-field magnets as used in magnetohydrodynamic generators, accelerators, nuclear fusion reactors, magnetic levitation, energy storage, medical instruments and so on. They could also be used in lossless power transmission lines. The Josephson effect of superconductors can be used to make ultra-fast computers and other electronic devices.

The highest T_c measured previously was 23 K, in a niobium-germanium alloy, which requires the use of very expensive liquid-helium coolant to achieve the superconducting state. Because the new materials superconduct at temperatures obtained by liquid nitrogen, which is much cheaper (22 cents per gallon rather than \$11), many applications are becoming feasible. As pointed out by Professor S. Nakajima (Tokai University) this discovery is comparable to, or even greater than, the invention of the transistor, because it affects not only the field of electronics but also the technology of high-power electricity. It is not an exaggeration to say that such a development will bring about a new industrial revolution, which is why so many people are so excited.

A special session was devoted to the superconductors at the annual meeting of the Physical Society of Japan on 28 March at Nagoya Institute of Technology, attended by more than 1,000 people (see page 432 of last week's issue). In addition to the 18 scheduled talks, 49 prepared discussion papers were accepted just before the session started. The session began at 9.30 a.m. and lasted until 10.40 p.m. The fevered atmosphere seems comparable with that at the American Physical Society Meeting held in New York on 18 March (see the article by M. Strongin *et al.* on pages 540–541 of last week's issue).

There are three classes of high- T_c ceramic superconductors (S. Tanaka, University of Tokyo): $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$ (a prototype of the newly discovered

compounds) with $T_c = 14$ K; La-X-Cu-O ($\text{X} = \text{Ba}, \text{Sr}, \text{Ca}$) for which $T_c \sim 40$ K; and Y-Ba-Cu-O for which $T_c \sim 90$ K. There are many reports claiming different T_c s in substances with different composition; but for them to be classed as real superconductors, Tanaka insisted on four conditions that had to be satisfied: the structure of the material should be clarified; the Meissner effect should be observed; the electrical resistivity should drop to zero; and the experiments should be reproducible.

The preparation of samples and the measurement of their electrical and magnetic properties were reported by many groups, including that of S. Hikama (University of Tokyo). This group discovered the second class of the new superconductors independently of C. W. Chu of the University of Houston. Agreement on their structure seems to have been obtained between the different groups. The La-X-Cu-O system has a K_2NiF_4 -type structure, and the Y-Ba-Cu-O system has the composition $1:2:3:7-\delta$. For the latter, however, there are still different models for the position of the oxygen atoms. Neutron diffraction reveals a model similar to that of the Bell group (F. Izumi, National Institute for Research on Inorganic Materials). For the first class of superconductors, the structural changes caused by temperature and the barium content x were examined: there are two tetragonal phases for large x (K. Ohbayashi, Hiroshima University).

The Hartree-Fock instability against spin density waves was examined theoretically as a function of x , and the phase diagram constructed (Y. Hasegawa, Institute of Solid State Physics). Such a phase diagram was determined experimentally by magnetic susceptibility measurements (T. Fujita, Hiroshima University). Band calculations were done by K. Takegahara *et al.* (Tohoku University) using the augmented plane-wave method. Ultraviolet photoelectron spectroscopy data show that the density of states at the Fermi level is rather small (T. Takahashi, Tohoku University). Surprisingly, little change is observed in T_c for the second class of superconductors, even if yttrium is substituted by a series of lanthanide atoms (H. Takagi, University of Tokyo). For the first class, the substitution does not destroy the superconductivity but T_c depends on the substituting lanthanide atom. A neutron-diffraction experiment reveals that there is a spin order in which each copper atom has a magnetic moment of $1.1 \mu_B$ (Yamaguchi, Tohoku University).

Besides sintered samples, various other

forms were also reported. M. Sato (Institute of Molecular Sciences) prepared single crystals and measured the anisotropy of the magnetoresistance. T. Murakami (Nippon Telephone and Telegraph Co.) prepared high-quality films by sputtering and characterized their electrical and magnetic properties. Kohno (Fujikura Co.) prepared wires in a german-silver tube without T_c being reduced compared with the sintered sample. Aoi (NEC Corp.) observed the Josephson effect and confirmed the singlet superconductivity from the duration of the switching steps.

Anomalous behaviour at room temperature was reported by K. Ohara (Kagoshima University). The a.c. and d.c. conductivities are different by two orders of magnitude, and the a.c. Josephson effect was observed at room temperature. But the state is very unstable, probably because of the non-bulk properties of the ceramic surfaces, and neither zero resistivity nor the Meissner effect was observed. K. Takita (Tsukuba University) also reported an unstable anomaly at 150 K as well as a strong dependence of the critical magnetic field (which destroys superconductivity) on temperature, $dH_{c2}/dT \sim 3 \text{ T K}^{-1}$. As well as the superconducting state, the physical properties of the materials in the normal states are also being studied. S. Uchida (University of Tokyo) said that his group is investigating the Hall effect, magnetization, plasma reflection and other properties.

The upper critical magnetic field H_{c2} and the critical current I_c are the two vital

parameters in applications of superconductors. H_{c2} was measured in pulsed high-magnetic fields at the Institute for Solid State Physics, Tokyo, and Uchida revealed that H_{c2} is higher than 50 T. A similar result was obtained by K. Okuda's group (Osaka University). Measurements in steady magnetic fields show that the onset temperature of superconducting transition is not changed much by magnetic fields, but the magnetoresistance increases appreciably at lower temperatures near T_c . It was pointed out that I_c is still small and insufficient for use in superconducting magnets. It is $3\text{--}4 \text{ A cm}^{-2}$ but does not depend strongly on magnetic field (K. Noto, Tohoku University).

As for the mechanism of the superconductivity, it remains unclear whether or not the materials are BCS-type superconductors with phonon-mediated Cooper pairs. Theoreticians proposed different models. A model of bipolaron tunnelling produced by the Jahn-Teller effect was proposed by H. Kamimura (University of Tokyo), and H. Fukuyama (Institute for Solid State Physics) calculated T_c by the coherent potential approximation based on the resonant valence-bond mechanism proposed by P. W. Anderson (*Science* **235**, 1196–1198; 1987). As Tanaka said, little was known of the properties of these materials before the sudden discovery of their superconducting phase, and we still do not know very much, so we must investigate them "with courage". □

Noboru Miura is at the Institute for Solid State Physics, University of Tokyo, Roppongi, Minato-ku, Tokyo 106, Japan.

Vertebrate palaeontology

Long night for owl monkeys

R. D. Martin

ALTHOUGH nocturnal life is typical of prosimian primates (lemurs, lorises and tarsiers), owl monkeys are the only nocturnally active representatives among the 'higher', simian primates (monkeys, apes and man). Until recently, however, there was no fossil evidence to document the antiquity of nocturnal habits of the owl monkey. Fragmentary new material recovered from middle Miocene deposits of La Venta, Colombia by Setoguchi and Rosenberger and described on page 692 of this issue¹ indicates that a large-eyed direct relative of the modern owl monkey was present in the South American fauna 12–15 million years ago. Indeed, the fossil form is so similar to the modern owl monkey in dental features that it has been allocated to the same genus, *Aotus*. On one hand, this indicates that the nocturnal habits of owl monkeys emerged considerably earlier than might otherwise have been supposed; on the other, it bears

witness to remarkable evolutionary conservatism in the owl monkey lineage.

As with other mammals, nocturnal habits among primates are generally correlated with relatively small body size and the owl monkey, with a body weight of about 1 kg, lies well within the body-size range of modern prosimians. However, there are other simian primates, notably the marmosets and tamarins, that are smaller than owl monkeys and yet are diurnal like other monkeys. It is therefore difficult to explain why owl monkeys, uniquely among the simians, should exhibit nocturnal activity. Perhaps, as there are no prosimian primates in the New World, owl monkeys have occupied a parallel ecological niche. It is interesting that owl monkeys, like prosimians, exhibit relatively low basal metabolic rates².

Several lines of evidence indicate that owl monkeys have undergone secondary adaptation for nocturnal habits. For



The modern owl monkey, *Aotus*, has the largest eyes relative to body size of any simian primate. (Photograph by Katharina Weiss, Anthropological Institute, Zürich.)

instance, they lack the reflecting tapetum found behind the retina of the eye in nocturnal strepsirrhine primates (lemurs and lorises) and they share with other simian primates the marked reduction in olfactory bulb volume that accompanied the original switch from nocturnal to diurnal habits in simian evolution³. Owl monkeys owe their name to the fact that, as an adaptation for vision in dim light conditions, the eyes have become secondarily enlarged (see figure), such that the relative size of the eyes is now comparable with that found among modern strepsirrhine primates. Nevertheless, owl monkeys do not seem to be so thoroughly adapted for night-time activity as nocturnal lemurs and lorises, as is indicated, for instance, by the fact that diurnal activity has been reported for owl monkeys in the extreme southern part of their range in Argentina⁴.

The fossil record of the New World monkeys is generally very poor. The earliest known fossil representative, documented exclusively by jaw fragments, is the Oligocene form *Branisella*, the reported age of which has recently been drastically reduced by almost 10 million years to 26–27 million years ago. Virtually all the remaining fossil New World monkeys come from late Oligocene–middle Miocene deposits and relatively little material has been recovered. Partial skulls are known only for the genera *Dolichocebus*, *Homunculus* and *Tremacebus* and, apart from scattered fragments, the postcranial skeleton is documented only for *Cebupithecia*. *Neosaimiri* and *Stirtonia* are known merely from jaws and teeth. A few intriguing subfossil specimens from central America were, until recently, all that remained to complete this short list of fossil New World monkeys.

To a large extent, such poor documen-

tation of the fossil history of New World monkeys is attributable to limited fossil-hunting activities, and renewed interest in this area has already begun to yield valuable results. The fossil owl monkey material, allocated to the species *Aotus dindensis*, was discovered during joint surveys by the Instituto Nacional de Investigaciones Geológico-Mineras, Colombia and Kyoto University. These surveys have yielded some other fragmentary specimens that may document the presence of as many as five additional genera⁵. These specimens include isolated teeth allocated to the genus *Kondous*, thought to be related to the modern spider monkeys⁶, and to the genus *Micodon*, believed (largely on grounds of small size) to be an early relative of the marmosets and tamarins⁷. New primate fossil material has also been discovered⁸ from late Oligocene deposits in Chubut Province, Argentina during a recent expedition conducted by the Museo Argentino de Ciencias Naturales, Buenos Aires.

The new material allocated to *A. dindensis* described in this issue¹ dramatically confirms recent suggestions that several modern lineages of modern New World monkeys became established at least during the early Miocene. In dental terms, at least, the Miocene *Neosaimiri* is very similar to the modern squirrel monkey (*Saimiri*), whereas *Stirtonia* closely resembles modern howler monkeys (*Alouatta*) and *Cebupithecia* may be linked to modern sakis (*Pithecia*). Setoguchi further suggests⁶ that the teeth of *Kondous* document early establishment of spider monkeys (*Ateles*). The material from *A. dindensis* is even more striking in that it seems to show that the modern genus *Aotus* can itself be traced back to at least 12 million years ago.

Accordingly, the general picture that is now emerging is that relatively little evolutionary change has taken place in New World monkeys since at least mid-Miocene times. But because the earliest known fossil New World monkeys date only from the late Oligocene, interpretation of the early origins of these primates is problematical. If it is assumed that the known fossil record represents a reliable sample of the occurrence of monkeys in the New World, the data of Setoguchi and Rosenberger indicate¹ that the various lineages of New World monkeys diverged very rapidly from some Oligocene common ancestor and then remained relatively conservative thereafter.

An alternative possibility, however, is that the New World monkeys originated far earlier than has often been inferred and that the early stages of their adaptive radiation have yet to be documented by fossil material. There is still a lively debate⁹ over the geographical origin of the common ancestor of the New World monkeys. Invasion from either North

America or Africa has been proposed. An early origin for the New World monkeys would fit well with the increasingly compelling evidence that the invaders came from Africa⁸, as progressive widening of the South Atlantic during the Tertiary renders any late invasion from Africa increasingly unlikely. □

R. D. Martin is director of the Anthropological Institute and Museum, University of Zürich, 8057 Zürich, Switzerland.

Quasicrystals

News on five-fold symmetry

P. Guyot

THE initial discovery¹ of 5-fold symmetry in aluminium-transition metal alloys was established from electron-diffraction patterns of materials of micrometre dimensions produced following rapid solidification. Soon afterwards, various techniques, ranging from melt-spinning, diffusion in the solid state and glass devitrification, were used to produce materials with point-group symmetries incompatible with crystal translational symmetry. Similar methods showed that a range of alloys tend to form quasicrystals: simple metals (Al-Mg-Zn, Al-Li-Cu); transition metals (Ni-Ti-V, V-Ni-Si); and rare earths or actinides (Co-Er, Pd-U-Si). Eventually, millimetre-sized single quasicrystals of $\text{Al}_6\text{Li}_3\text{Cu}$ were produced² by slow solidification. These quasicrystals are thought to be in a stable equilibrium state with a well-faceted external morphology — a rhombic triacontahedron — consistent with local atomic units proposed in modelling the icosahedral structure. The aim of a recent meeting* was to analyse new experimental results that bear on the atomic structure, growth, morphology and other properties of quasicrystals.

There are two classes of models of the structure of quasicrystals that seem to be currently favoured: three-dimensional Penrose tiling, either perfect or defected by phasons or dislocations; and packing of strongly bound icosahedral units in the same orientation with a high degree of connection randomness (the icosahedral glass).

Figure 1a shows a spectacular aggregate found by J.M. Lang and collaborators (Centre de Recherches Cédur-Péchiney, Voreppe)³ in slowly cast $\text{Al}_6\text{Li}_3\text{Cu}$: this structure is composed of 12 triacontahedra whose centres lie on the vertices of an icosahedron seen along a 5-fold axis (see also the front cover of this issue). This geometry implies an underlying central 20-branched star (Fig. 1b). Such a star can be considered to be multi-

1. Setoguchi, T. & Rosenberger, A.L. *Nature* **326**, 692–694 (1987).
2. Müller, E.F. *Bonn. zool. Beitr.* **34**, 29–71 (1983).
3. Martin, R.D. in *The Study of Prosimian Behavior* (eds Doyle, G.A. & Martin, R.D.) 45–77 (Academic, New York, 1979).
4. Rathbun, G.B. & Gache, M. *Primates* **21**, 211–219 (1980).
5. Setoguchi, T. & Rosenberger, A.L. *J. vert. Paleont.* (in the press).
6. Setoguchi, T. *Folia primat.* **44**, 96–101 (1985).
7. Setoguchi, T. & Rosenberger, A.L. *Int. J. Primat.* **6**, 615–625 (1985).
8. Fleagle, J.G. & Bown, T.M. *Folia primat.* **41**, 240–266 (1983).
9. Ciochon, R.L. & Chiarelli, A.B. (eds) *Evolutionary Biology of the New World Monkeys and Continental Drift* (Plenum, New York, 1980).

twins of a rhombohedral crystal. If large-unit-cell crystals are recognized to be good approximations to the ideal quasicrystal, twinning, in view of the results obtained using dark-field electron microscopy, seems more likely to be involved in the growth process rather than to constitute a basic structural ingredient.

But new results on diffraction patterns of single quasicrystals of $\text{Al}_6\text{Li}_3\text{Cu}$, demonstrate that the structure is still far from fully understood: the monochromatic transmission Laue X-ray patterns presented by F. Dénoyer and colleagues⁴ (shown in Fig. 2 for the 5-fold orientation) confirm the icosahedral symmetry but reveal a high density of diffraction spots, and also complex diffuse scattering. As well as neutron-scattering experiments based on contrast variation by isomorphous substitution on the transition metal

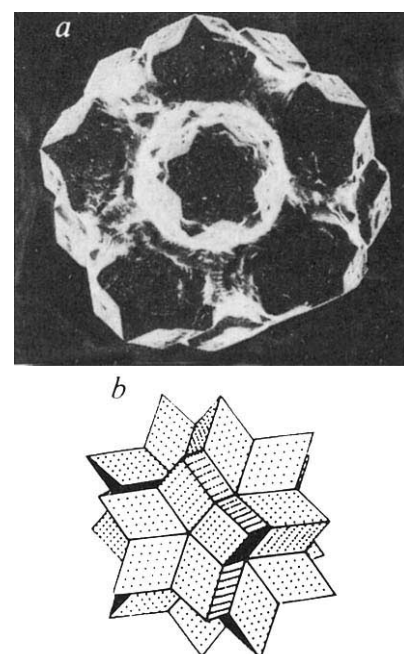


Fig. 1 a, Quasicrystal aggregate of $\text{Al}_6\text{Li}_3\text{Cu}$; courtesy of Dr B. Dubost. b, Central 20-branched star inferred from structure of a.

*Workshop on Quasicrystals, Grenoble, 22–23 January 1987.

sites of aluminium-transition metal quasicrystals, in order to separate the contributions to the diffraction pattern from the sublattices, Janot *et al.* (Laue-Langevin Institute, Grenoble) also reported the existence of strong diffuse scattering in between the Bragg peaks. Diffuse scattering and peak broadening obviously need to be studied in detail before they can be tentatively correlated with local atomic order or defects such as phasons or wrong connections. These parameters also compromise the significance of the reconstructed Patterson function in periodic six-dimensional space, analysis of which, according to D. Gratias (Centre d'Etudes de Chimie Métallurgique, Vitry), already suffers intrinsically from a degree of freedom on the phases and therefore can only lead to a class of locally isomorphic structures.

From the current state of knowledge

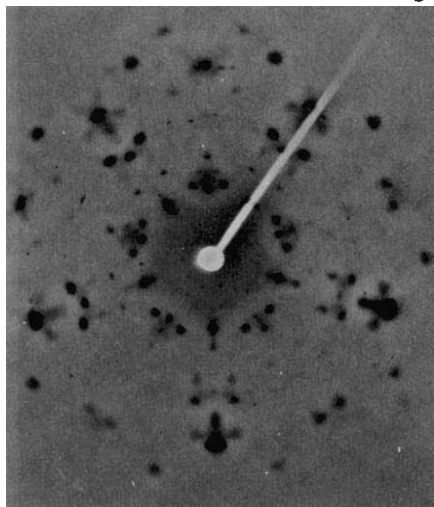


Fig. 2 Laue X-ray diffraction pattern of an $\text{Al}_6\text{Li}_7\text{Cu}$ quasicrystal; courtesy of Professor F. Dénoyer.

about the atomic structure, physical properties of real quasicrystals have been predicted to lie between those of crystals and metallic glasses. Indeed, this trend seems to be observed within the limits of well-defined specimens. For a 'macroscopic' disorder, the concentration gradients, multi-phase states and various types of interface that exist in melt-spun materials (J.L. Verger-Gaugry *et al.*, Lab. de Thermodynamique et Physico-Chimie Métallurgiques, Grenoble) may be the cause of contradictions between some recently reported measurements.

J.B. Suck (Kernforschungszentrum, Karlsruhe), using inelastic neutron scattering, found that the atomic dynamics of icosahedral $\text{Pd}_{58.8}\text{Si}_{20.6}\text{U}_{20.6}$ are very similar to those of the metallic glass of the same composition, both having a generalized vibrational density of states different from that of the fully crystallized state. Quasicrystals and glass share low-energy modes, with the same dependence on momentum transfer. The vibrational part of the specific heat calculated from the density of

In a report last year, J. Marks, J.-P. Shaw and C.-K. J. Shen (*Nature* 321, 785; 1986) added a new member to the α -globin gene family. However, it was unclear at the time whether the gene encoded a subunit of a hitherto unknown and unsuspected form of haemoglobin — perhaps one specialized for carrying oxygen at a very early stage of development — or whether, as intimated by N. Proudfoot in a News and Views article (*Nature* 321, 730; 1986) it was merely another pseudogene, one of those ghostly shadows of real genes caught at various stages of decay into randomness. The doubts about the new gene, dubbed a θ 1-globin gene, were stimulated by the unusual nature of the sequence flanking the presumed site of initiation of transcription, where signals regulating the expression of the gene are expected to be found. Although the CCAAT and ATA sequences characteristic of eukaryotic 'promoters' were identified, they were considerably further upstream of the coding sequence of the gene than usual, perhaps, as Proudfoot suggested, too far for the gene to be expressed.

Although direct evidence that the θ 1-globin gene is functional in any species is still lacking, the results reported on pages 717–720 of this issue considerably strengthen the case. Marks's group has now cloned and sequenced the equivalent to the orang-utan θ 1-globin gene from the olive baboon. The availability of the two

states is also in good agreement with the specific heat measured at low temperatures for the metallic glass and the quasicrystal. Similarly, the local moments appearing on the manganese atoms of $\text{Al}_{73}\text{Mn}_{21}\text{Si}_6$, deduced from the comparison of low-field and high-field magnetization measurements (R. Bellissuet *et al.*, Léon Brillouin Institute, Saclay), interact magnetically below 5.2 K in the icosahedral phase with a characteristic spin-glass freezing, as previously suggested⁵ by workers at Bell Laboratories in the amorphous state. The magnetism, which disappears in the crystalline β -phase, could possibly result from a proximity effect of nearest-neighbour manganese atoms.

Heat capacity, transport and optical properties of Al–Mn–(Si) were also reported by C. Berger *et al.* (Lab. d'Etudes des Propriétés Electroniques des Solides, Centre de Recherches sur les Très Basses Températures, Grenoble), J.L. Verger-Gaugry *et al.* (Lab. de Thermodynamique et Physico-Chimie Métallurgiques, Grenoble) and J.M. Frigerio *et al.* (Lab. Optique des Solides, Orsay). Clearly, these properties appear to be strongly influenced by scattering on manganese d resonant states in the aluminium sp conduction band. Whereas the high

sequences allows an evolutionary test of whether the θ 1-globin gene is functional in these species. This is because the sequence of a gene can be partitioned into a subset of sites expected to be subject to conservative selective pressure (purifying selection) if the gene is functional, and a subset expected to evolve more rapidly by neutral drift. The expectation is that the difference in rates of evolution of these two subsets should be lost if the gene is a pseudogene. When Marks and colleagues compared the sequences of the orang-utan and olive baboon θ 1-globin genes they found a clear difference in the amount of divergence of the putative selected and non-selected subsets of the sequence: of 27 changes in the putative coding region of the genes, 22 are in the silent third codon positions and only 5 are predicted to alter the sequence of the encoded protein. In fact the third codon positions of this gene seem to be evolving remarkably quickly, so the result is particularly clear-cut. Taken together with the conservation of the unusual promoter organization in the olive baboon θ 1-globin gene, the balance of probability must surely now lie with the gene being functional. If this is indeed so it will be of interest to learn precisely when during development the gene is expressed, and what features the haemoglobin containing the θ 1-subunit may have that suit it to carrying oxygen during that developmental stage.

Geoffrey North

values of the direct current resistivity are of the same order of magnitude as in the amorphous phases, indicating a contribution from structural defects, the frequency-dependent conductivity is similar to that in crystalline compounds, presumably with band structure-like effects.

Finally, Pannetier *et al.* (Centre de Recherches sur les Très Basses Températures, Grenoble) demonstrated the quantization of magnetic flux across a superconducting micronic mesh of aluminium or indium wires reproducing a Penrose tiling, similar to the work⁶ of a group at the University of Pennsylvania. Access to the singular electronic states of such a perfect two-dimensional quasicrystal can soon be expected. But transposition to real alloys is probably more hazardous. □

1. Shechtman, D., Blech, I., Gratias, D. & Cahn, J.W. *Phys. Rev. Lett.* 53, 1951 (1984).
2. Dubost, B., Lang, J.M., Tanaka, M., Sainfort, P. & Audier, M. *Nature* 324, 48 (1986).
3. Lang, J.M., Audier, M., Dubost, B. & Sainfort, P. *J. Cryst. Growth* (in the press).
4. Dénoyer, F., Heger, G., Lambert, M. & Lang, J.M. *C.R. Acad. Sc. Seance* (in the press).
5. Hauser, J.J., Chen, H.S. & Waszczak, J.V. *Phys. Rev.* B33, 3577 (1986).
6. Behrooz, A. *et al. Phys. Rev. Lett.* 57, 368 (1986).

P. Guyot is at the Institute National Polytechnique de Grenoble, Domain Universitaire, B.P.75, 38400 St-Martin-D'Hères, Cedex, France.

Conservation biology

Living Latin binomials

Robert M. May and Anna Marie Lyles

RISE on wooded terraces above Sydney Harbour, Taronga Park is surely the world's most beautifully situated zoo. It was also a pioneer in housing many animals in 'naturalistic' settings rather than in barred cells. One of us has early memories of seeing lions in a large sunken arena, with gothic grottos around the edges and a raised central platform (in the shape of a map of Australia!) on which the lion pride usually lazed. Such an exhibit clearly provides an evocative display of the Latin binomial *Leo leo* in living form. But the extent to which these creatures represent the African lion, with the rich array of individual and group behaviour that is increasingly being revealed^{1,2} is another question. Such questions — heretofore largely philosophical — have recently taken on a new urgency, as a result of efforts to reintroduce zoo-bred animals into their original habitats in the wild (see our News and Views article³ last month and ref. 4).

The difficulties inherent in such reintroductions of animals which have flexible, learned behaviour patterns are well-illustrated by the golden-lion tamarin, *Leontopithecus rosalia*⁵ (see figure). These tiny primates are covered in long golden-orange fur (like an extended mane, hence the 'golden lion'); with their inquisitive, engaging behaviour and their twittering bird-like song, they seem designed to tug at the heart-strings. The golden-lion tamarin is native to southeastern Brazil, where the loss of an estimated 98 per cent of their habitat, coupled with exploitation in the 1960s for the pet and zoo trade, has brought the species to the brink of extinction in the wild. Over the past few decades, some habitat has been protected and captive populations have grown from around 80 to 430 individuals (partly from additional captures, partly from reproduction in captivity). The major wild population, located in the Poco das Antas Reserve, recently numbered only about 100 tamarins, and their blood proteins seem to exhibit less genetic diversity than those of the captive tamarins. The recent programme of reintroducing captive tamarins in Poco das Antas is partly motivated by the hope of ameliorating inbreeding problems by exchanging genes between wild and captive populations.

Of the 26 golden-lion tamarins reintroduced in Brazil during the past two years or so, five are known to be alive, three are

missing and the others are dead. Most deaths occurred shortly after release and were caused by disease, predation, exposure, starvation, snakebite and social conflict. This hefty toll occurred despite efforts to acclimatize and train these animals for life outside captivity.

Before being released, the tamarins were quarantined, acclimatized to novel, more natural foods, and trained to hunt for food that was dispersed and hidden in their cages. After release, the tamarins

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Group of golden-lion tamarins in captivity before going to Brazil (photograph by Jesse Cohen, National Zoological Park, Washington DC).

seemed to avoid some potentially noxious foods like mushrooms, but they showed little selectivity towards others. Several animals became seriously ill after attempting to eat a toad, but they nevertheless persisted in their attempts the next day. They also tried to mob and eat snakes; one died on one occasion when the snake was poisonous.

Attempts were made to prepare the tamarins for jumping, climbing and hanging on supports that, like natural vegetation and unlike the rigid structures in most cages, were not fixed in place. The tamarins, however, turned out to be reluctant to use natural substrates, and after being released they kept returning to their cages which eventually had to be dismantled to force the animals to enter the trees. To compound these difficulties, the tamarins

seemed incapable of making cognitive maps of the forest; some became lost and disoriented, while others simply sat on the ground and were killed by predators.

Disease was the main cause of mortality among the reintroduced tamarins. The reasons are not well understood, and the subject seems to have received less attention than the more directly observable factors influencing survival that we have discussed here. It may be, in part, that levels of herd immunity decrease in captivity (as a result of smaller population sizes), leaving adults susceptible to infections that would be less lethal if experienced in early years (paralytic poliomyelitis among humans may be an analogy)⁶. Lower levels of herd immunity also imply that fewer infants will be protected for the

first few months of life by maternally derived antibodies (because such maternal antibodies are present only if the mother has at some time experienced infection). This is all conjecture, but the fact remains that disease is a leading killer of reintroduced tamarins. Some kind of immunological preparation, possibly including vaccination against specific infections, may come to be a significant part of reintroduction programmes.

The initial attempts at reintroducing tamarins have led to refined protocols for training. In future, young tamarins will be preferred for reintroduction because they are more successful than older ones (perhaps because of greater flexibility in their behaviour). Interestingly, three of the known five survivors in Poco das Antas live in two groups with wild-born mates, and it seems likely to us that the mates are easing the process of readjustment. Encouragingly, the survivors have been breeding, and in this sense the programme is successful. In considering the golden-lion tamarin programme, we should

also bear in mind the complexities and dangers that exist for small creatures such as these in a tropical forest; most other reintroduction programmes are in simpler, less precarious environments.

One of the most ambitious schemes for reintroduction is Yalden's proposal⁷ to recreate on the Scottish island of Rhum the large mammal fauna of western Europe as a tourist attraction. This would involve reintroduction of animals, such as the polecat, pine marten and wolf, that were exterminated from Britain in the past century. It also envisions reintroduction of some extinct species, such as aurochs and tarpan, that have been genetically 'reconstituted' from domestic descendants. This proposal seems to us to have given careful thought to the genetic technology, and some thought to habitat

questions (Rhum vegetation today is not exactly what it was before the advent of humans); but it seems to have given little consideration to the behaviour of reconstituted aurochs and other such beasts, or to the effects of this behaviour on the functioning and persistence of the proposed ecosystems. As long as what is proposed is a kind of upmarket Disneyland, these worries may be insubstantial. Serious attempts to preserve biological species must, however, deal with vexing questions of preserving not just the species but also its pristine range of behaviour.

In summary, there is a difference between *L. rosalia* securely preserved in the zoo and the golden-lion tamarin with the rich spectrum of learned behaviour neces-

sary for survival in the wild. This difference is more than a topic for philosophical debate; in the long run, such 'cultural' losses may be the most significant problem when and if we try to unload the ark. □

1. Schaller, G.B. *The Serengeti Lion* (University of Chicago Press, 1972).
2. Packer C. in *Ecological Aspects of Social Evolution* (eds Rubenstein, D.I. & Wrangham R.W.) 429-451 (Princeton University Press, 1986).
3. Lyles, A.M. & May, R.M. *Nature* **326**, 245-246 (1987).
4. Palca, J. *Nature* **323**, 7 (1986).
5. Kleiman, D.V. et al. in *Primates: The Road to Self-Sustaining Populations* (ed. Benirschke, K.) 959-979 (Springer, New York, 1986).
6. Anderson, R.M. & May, R.M. *Nature* **318**, 323-329 (1985).
7. Yalden, D.W. *Mammal Rev.* **16**, 53-63 (1986).

Anna Marie Lyles and Robert M. May are in the Biology Department, Princeton University, Princeton, New Jersey 08544, USA.

Radioastronomy

Spectra of stellar radio flares

Arnold O. Benz

THE radio emission of stellar flares has attracted increasing attention in recent years as a result of instrumental advances. Red dwarf (dMe) stars are known to be magnetically active and variable on time-scales of minutes (and less) in a wide range of emissions including X-rays, optical and radio waves. Now an important development has started with the first spectral measurement, reported by Bastian and Bookbinder on page 678 of this issue¹. They used the Very Large Array telescope in New Mexico (currently the world's largest radio interferometer) at 21-cm wavelength to record the radio spectra of two outbursts of the flare star UV Ceti (L726-8B). The range of frequencies (bandwidth) accepted by this instrument is rather limited, but is sufficient to show irregular variations in the spectrum of one event. The results show a potential to measure electron density and magnetic field strength in the atmospheres of dMe stars. Various plasma processes seem to be at work which first have to be disentangled. The door is not yet really open, but the glance through the keyhole looks promising.

The situation of the radio observations of flare stars is reminiscent of solar radioastronomy in the early 1950s when Wild² introduced radio spectrogrammes, on which the variation of intensity with time as well as radio frequency is shown. This method added a new dimension and unravelled structures which were unrecognized in single-frequency recordings. It quickly led to a number of important discoveries, such as mildly relativistic particle beams (type III bursts); coronal shock waves (type II bursts); and large magnetic loops interconnecting remote active regions on the Sun (type U-bursts).

Later, spectral measurements at micro-

wave frequencies revealed that common solar flares accelerate a large number of relativistic electrons which spiral around magnetic field lines and emit synchrotron radiation. In the following two decades, observations in the low-frequency part of the spectrum brought insights into plasma processes in the outer corona and in interplanetary space. Recently, millisecond spikes with narrow bandwidth have been suggested to be intimately associated with the flare-energy release proper. It is remarkable that all these findings were made by full Sun observations without spatial resolution, just as will be the case for stars.

Flare stars were the first stars after the Sun³ detected in the radio frequency. They are very abundant, comprising about 10 per cent of all stars in our Galaxy, but only about a dozen, the nearest specimens, are well observed. Early reports were plagued by confusion of terrestrial interference with stellar events. Reliable measurements came with the use of large telescopes such as Arecibo, and of interferometers. Such observations have, for example, revealed a quiescent radio emission⁴ of UV Ceti which is 100 times more powerful than that of the Sun, and a slowly varying component¹ even larger. Stellar radio flares, observed at a single frequency, are very frequent and up to 10⁵ times higher in luminosity than solar counterparts. It is generally agreed that these flares are a manifestation of a strong magnetic field built up by vigorous convection in late-type, cool and small stars. The highly polarized flare radio emission and the recent discovery of millisecond structures⁵ in a similar star strongly indicate that such sources are much smaller than their associated star and have a brightness temperature of at least 10¹⁸ K.

For this to be so, the emission process must be coherent, which might imply that maser-type action occurs in the cyclotron radiation (generated by non-relativistic electrons). That is to say, an electron-cyclotron instability arises because of the distribution of electron spiralling velocities being peaked to the higher velocities — a population inversion.

The new spectral observations¹ independently confirm the small source size. In one event spectral components with width approximately 0.2 per cent of their central frequency were discovered. Such narrow-band emission is only possible in a correspondingly homogeneous source of much smaller dimension than the scale size, because the variation of the parameter that determines the emission frequency (such as the cyclotron frequency) must be within a small margin. This requirement puts an upper limit on the source diameter of 200 km, assuming a scale height of one stellar radius, the largest reasonable. The above emission mechanism implies a magnetic field in the source of 250 gauss (assuming the radio frequency is the second harmonic of the cyclotron frequency, which would be usual) and an upper limit on the electron density of about 10⁹ cm⁻³, which is imposed if the maser is to operate.

It is certainly tempting to associate the observed stellar events with solar-burst types and processes, but the necessary resolution, simultaneously in time and frequency, is not yet available. Narrow-band components have been observed in solar radio spikes, but also as fine structure of decimetric type IV (due to trapped flare particles) and type II bursts. Fast structure in time is present in solar decimetric pulsations, type I bursts (the result of new magnetic flux penetrating the corona) as well as spikes. Broad-band observations have been found to be necessary to assess solar radio bursts. Clearly a time-resolution better than one second is another requirement.

More importantly, the lesson of solar astronomy is that an understanding of turbulent plasmas and particle beams will not come from radio observations alone. The same excitors also produce signatures in hard X-ray and ultraviolet lines. Stellar flares may contribute significantly to the universal soft X-ray background. Finally, radio bursts may reveal the geometry of the powerful magnetic field and related processes in the atmosphere of the probably most abundant variable stars. □

1. Bastian, T.S. & Bookbinder, J.A. *Nature* **326**, 678-680 (1987).
2. Wild, J.P. & McCready, L.L. *Aust. J. Sci. Res.* **A3**, 387 (1950).
3. Lovell, B. *Observatory* **84**, 191 (1964).
4. Gary, D.E. & Linsky, J.L. *Astrophys. J.* **250**, 284 (1981).
5. Lang, K.R. & Willson, R.F. *Astrophys. J.* **305**, 363 (1986).

Arnold O. Benz is at the Institute of Astronomy, Eidgenössische Technische Hochschule, 8092 Zürich, Switzerland.

Atrial natriuretic peptide

How the heart rules the kidneys

Brenda Leckie

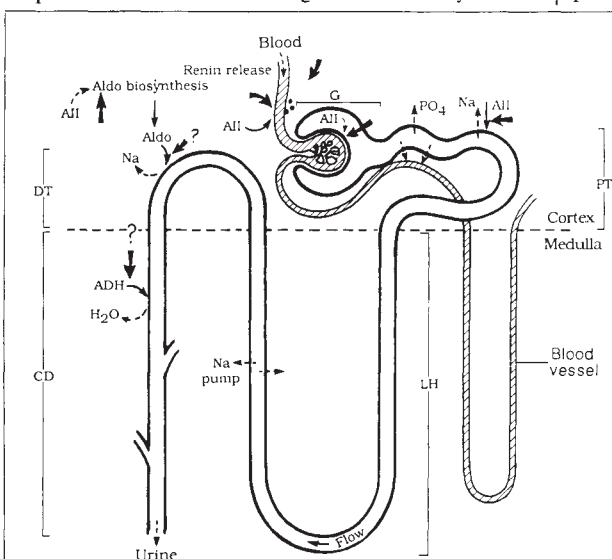
ATRIAL natriuretic peptide (ANP), secreted by the heart, increases salt and water output by the kidney. On page 697 of this issue¹, Harris, Thomas and Morgan report the use of an elegant microperefusion technique to show that this peptide blocks the uptake of sodium and water passing through the renal tubule by inhibiting the effect of angiotensin II, a hormone that promotes this reabsorption. These results provide new insights into the renal action of ANP.

Production of urine by the kidney may seem a mundane topic, but it is a complex process that is still not completely understood. Blood flows through the glomeruli which filter water, salts and other low-molecular-mass solutes into the renal tubules (see figure). The glomerular filtration rate is about 120 millilitres per minute; thus the whole blood volume of around 7 litres is filtered each hour — a rate that would clearly be incompatible with life unless the volume of urine was reduced before it was voided. The ultra-filtrate from the glomeruli is concentrated in the renal tubules where useful components such as glucose and protein are reabsorbed, waste products are excreted and the amount of salt and water is adjusted so that output in the urine balances intake.

When ANP is infused into normal humans, there is a dramatic increase in urine volume, in the output of sodium, magnesium, calcium and phosphate, and sometimes in the output of potassium. How does this happen? ANP always increases the glomerular filtration rate and the amount of filtrate produced from each millilitre of blood passing through the kidney. This rise in filtration fraction could occur because the blood vessels supplying the glomerulus are dilated by ANP, whereas vessels draining the glomerulus are unaffected or even constricted. Such an effect has been reported recently². ANP seems to relax smooth muscle, particularly arteries in the kidney, by blocking vasoconstrictors such as noradrenaline and angiotensin II. The relaxation of the incoming (afferent) glomerular arteriole may be caused by the peptide blocking the action of angiotensin II. The concentration of angiotensin II is likely to be particularly high in this arteriole because the enzymes that produce it, renin and converting enzyme, are

also present there.

The glomeruli contain high-affinity receptors for ANP³. On its own, ANP does not affect glomerular size, but it does inhibit the contractile effect of angiotensin II on glomerular mesangial cells, which increases the surface area of the glomerulus and presumably the flow of filtrate through it⁴. The fact that ANP increases the filtration through single glomeruli could be important for the treatment of types of renal failure where some glomeruli are blocked but others are normal. Increasing the efficiency of the



The filtration apparatus of the kidney, showing the sites of action of ANP. G, glomerulus; PT, proximal tubule; LH, loop of Henle; DT, distal tubule; CD, collecting duct; Aldo, aldosterone; AII, angiotensin II. Heavy arrows, sites of ANP action; light arrows, action of other hormones; dotted arrows, salt and water movement.

normal nephrons might improve overall renal function. The effect of ANP on both the renal vasculature and the glomerular cells would be to increase the flow of urine, which might also be caused by washing out the osmotic gradient between cortex and medulla that concentrates the urine.

Some people think that the natriuretic and diuretic effects of ANP are caused entirely by its action on the renal blood vessels, but others believe that the peptide alters tubular function. Clearance studies, in which the kidney is considered as a 'black box' and measurements are made of ionic concentrations entering in the blood and leaving in the urine, provide conflicting evidence, as do studies on sodium transport in isolated tubules. The results of Harris *et al.* reported in this issue¹ demonstrate that ANP reduces proximal tubular reabsorption by antag-

onizing the effect of angiotensin II infused directly into the peritubular capillary. The mechanism is unknown, not least because proximal-tubule sodium reabsorption is still obscure. Interference could be at the level of active sodium transport through the tubular cell, or with the entry of sodium into the peritubular capillary. Sodium entry is influenced by changes in capillary fluid pressure and angiotensin II could theoretically affect this. One slightly puzzling aspect is that only low-affinity receptors for ANP have been found on the proximal tubule compared with other parts of the kidney. Maybe these are effective, or perhaps a small population of high-affinity receptors exists either on the tubule itself or on the peritubular capillaries. Because 80 per cent of filtered sodium and water is reabsorbed in the proximal tubule a relatively small decrease in absorption could lead to a large change in urine output.

What other effects could ANP have on the kidney? Receptors for this peptide are present not only in the blood vessels and glomeruli but also in the distal tubule and collecting ducts. ANP could block the absorption of sodium from the distal tubule by opposing the action of aldosterone, which promotes absorption at this site. Such a direct blocking effect has not been demonstrated, but ANP does reduce aldosterone synthesis by the adrenal gland⁵. Where aldosterone synthesis is high, as in the case of people on a low-salt diet, this suppressive effect of ANP might be small and distal tubular sodium reabsorption, with the associated outward passage of potassium into the tubule, could still occur while ANP 'upstream' decreases proximal tubular sodium reabsorption. The passage of the extra sodium from the proximal into the distal tubule could augment the distal sodium-potassium exchange even more to result in a significant potassium loss. This mechanism contrasts with situations where distal tubule sodium reabsorption is less active. This could be important in therapy, where it might be desirable to conserve potassium.

ANP could oppose the action of other hormones, such as calcitonin and parathyroid hormone, and thus influence calcium and magnesium excretion. Anti-diuretic hormone enhances water absorption from the distal tubule and if ANP opposes this action, increased flow (water diuresis) would result. The importance of antidiuretic hormone is shown by diseases in which it lacking, where large volumes of dilute urine are produced. If ANP acts mainly by antagonizing other hormones, it will display various effects depending on

the hormonal status of the target organ. The classical idea of a specific chemical transmitter may not be appropriate.

The many effects of ANP on the kidney will make sense only when the actions of the hormone at a cellular and molecular level are understood. Binding of ANP to its receptor causes release of cyclic GMP⁶ which alters calcium flux and cyclic AMP levels in some cells⁷. In the kidney, ANP increases cyclic GMP concentration in glomeruli and distal tubules. By itself it does not alter cyclic AMP levels, but it inhibits the forskolin- and antidiuretic hormone-induced rise in cyclic AMP⁸. Lowering of intracellular cyclic AMP does not explain why ANP blocks the action of angiotensin II, which could occur through the inhibition of receptor-mediated calcium flux or calcium mobilization. Finally, although ANP may control other hormones at the intracellular level, it may also have a direct action on cell function.

The renal effects of administered ANP are likely to vary with electrolyte and hormonal status of the subject. People with high circulating concentrations of angiotensin II will probably show an exaggeration of the angiotensin-blocking actions of ANP, whereas those with high levels of antidiuretic hormone might show a response related to blockage of this hormone. Variations of response could

occur in different diseases; for instance, hypertensives lose more water and salt than normal people when infused with ANP⁹. Some apparent contradictions in the literature on the renal effects of ANP probably result from the extent to which its actions are modified by other factors.

Knowledge of the renal actions of ANP will have practical consequences. The unique vasodilator and diuretic action of the hormone suggests that analogues could be useful in the treatment of heart failure, hypertension and renal failure. It might be possible to design compounds that act selectively on the blood vessels or on the tubules of the kidney. The study of ANP is advancing with remarkable speed and the momentum shows little sign of slowing down. □

1. Harris, P.J., Thomas, D. & Morgan, T.O. *Nature* **326**, 697–698 (1987).
2. Marin-Grez, M., Fleming, J.T. & Steinhausen, M. *Nature* **324**, 473–476 (1986).
3. Bianchi, C. et al. *Am. J. Physiol.* **251**, F594–F602 (1986).
4. Appel, R.G. et al. *Am. J. Physiol.* **251**, F1036–F1042 (1986).
5. Atarashi, K. et al. *Science* **224**, 992–994 (1984).
6. Ohlstein, E.H. & Berkowitz, B.A. *Hypertension* **7**, 306–310 (1985).
7. Hartzell, H.C. & Fischmeister, R. *Nature* **323**, 273–275 (1986).
8. Ishikawa, S. et al. *Biochem. biophys. Res. Commun.* **130**, 1147–1153 (1985).
9. Richards, A.M. et al. *Hypertension* **7**, 812–817 (1985).

Brenda Leckie is in the MRC Blood Pressure Unit, Western Infirmary, Glasgow G11 6NT, UK.

Palaeontology

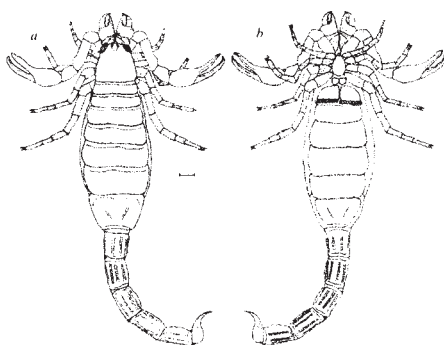
Scorpions take to the water

D.E.G. Briggs

SCORPIONS form an insignificant proportion of the living biota, exclusively terrestrial and celebrated only because of their infamous sting. Kjellesvig-Waering's mammoth new account of the fossil Scorpionida of the world (*Palaeontographica Americana* **55**, 1–287; 1986) corrects our currently distorted view of extinct scorpions as low in diversity and mainly air-breathing. Most scorpions in fact respired with gills, aquatic forms becoming extinct only after the Triassic. New evidence indicates that they are a sister group of the eurypterids, and confirms that the traditional division of chelicerates into the mainly aquatic Merostomata (xiphosurids and eurypterids) and the mainly terrestrial Arachnida (including scorpions, spiders and mites) is an artificial one based largely on environmental preference.

Although most fossil scorpions were aquatic, like other chelicerates they lacked a mineralized exoskeleton, and therefore require exceptional conditions for preservation. The earliest scorpions occur in middle Silurian rocks but their initial morphological diversity suggests that they are even older. Kjellesvig-Waering's monograph almost doubles the number of fossil

genera, documenting more than 30 new genera and species. His work spanned a period of about 20 years and, although it was largely complete on his death in 1979, the material was in a chaotic state. Its publication owes much to the painstaking labour of his colleagues Kenneth and Anneliese Caster of Cincinnati.



Proscorpius, an early scorpion from the Upper Silurian Bertie Waterlime of New York State. a, Dorsal side; b, ventral side showing five abdominal plates (the first partially concealed by pectines) that cover the gills and correspond in number and arrangement to those of eurypterids. (Reconstruction from the new work of E.N. Kjellesvig-Waering.)

Living scorpions respire through apertures (stigmata) in the abdominal sternites, which lead to book-lungs. These pulmonate scorpions first appeared in the late Carboniferous. The abdominal structure of all fossil scorpions were previously also interpreted, by analogy, as sternites. A group of large Palaeozoic chelicerates, the eurypterids, had gill tracts on the surface of the sternites, but they were covered by abdominal plates which formed a cavity opening posteriorly. Kjellesvig-Waering recognizes that the features in many fossil scorpions previously interpreted as sternites (by analogy with extant air-breathing forms) are in fact abdominal plates like those of eurypterids. The cuticle of the respiratory structures concealed by them was naturally very delicate and the chances of preservation are minimal. Remarkably, Kjellesvig-Waering describes the first known examples of such structures in Palaeozoic scorpions, represented by cuticle fragments. Those of *Tiphoscorpio* from the Devonian of New York State consist of discrete lamellae and seem to represent true gills.

The other well-preserved respiratory cuticle, that of *Paraisobuthus* from the Carboniferous of Nottinghamshire, is different, consisting of a spongy surface bearing cones and spines. This structure is similar to the gill tracts in some eurypterids. Selden (*Phil. Trans. R. Soc. B* **309**, 219; 1984) argued that the tracts in eurypterids are too small to have functioned in aquatic respiration and that they may have been only accessory respiratory organs for breathing air; in that case normal aquatic respiration must have been carried out by true gills which have yet to be discovered. This may also have been the case in *Paraisobuthus* and other fossil scorpions; the tract may have allowed it to adopt an amphibious lifestyle in the Coal Measure swamps, and true lamellate gills may also have been present. Kjellesvig-Waering considers that the critical transition from gills to lungs in scorpions involved the reduction and eventual loss of the abdominal plates. The sternites below became strengthened and developed perforations leading to the book-lungs.

The interpretation of fossil scorpions as aquatic does not rely solely on the discovery of gills. The sedimentology and associated faunas provide good evidence that some genera, such as *Palaeoscorpion* and *Palaeophonus*, were even marine. Certain structures present in living scorpions must have functioned rather differently in an aquatic medium. The position of the pectines, for example, which also occur in the Palaeozoic forms, may indicate some role in reproduction. They may have been used originally for scooping a hollow into which eggs were laid before fertilization by the male, rather in the manner of the horseshoe crab *Limulus*.

Some fossil scorpions reached substantial sizes. The identity of the Lower Devonian *Praearcturus gigas* is finally confirmed as a scorpion, making it the largest known species, reaching lengths of 1 metre. Although this species, together with *Brontoscorpia*, which approached the same dimensions, occur in terrestrial sediments, they are too large to have moulted on land, and must have been amphibious if not aquatic. The earliest unequivocal terrestrial scorpion is *Palaeopisthacanthus* from the Upper Carboniferous Mazon Creek biota of Illinois in which stigmata are preserved (Rolfe, W.D.I. *Spec. Vol. Syst. Assn* 15, 146; 1980). Carboniferous coals yield much scorpion material, and the unique hyaline layer in scorpion cuticle may enhance its

preservation potential compared with that of other arthropods (Bartram *et al.* *J. geol. Soc. Lond.* 144, in the press).

Kjellesvig-Waering's work reveals striking similarities between the preabdominal morphology and respiratory structures of aquatic scorpions and those of eurypterids, and indicates that these two groups are closely related. This supports recent classifications that no longer group the eurypterids with the xiphosurids in the class Merostomata, but recognize that the Xiphosura is the primitive sister group of other chelicerates (for example, Weygoldt, P. & Paulus, H.F. *Z. Zool. Syst. Evolut.-forsch.* 17, 177; 1979). □

D. E. G. Briggs is in the Department of Geology, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.

Cell biology

Synapsin I and the cytoskeleton

A. J. Baines

IN THE search for clues to the mechanism of nervous transmission, proteins associated with structures at synaptic termini are being probed in ever-increasing detail. At the focus of much effort is synapsin I, a protein associated with small synaptic vesicles. Recent results from several groups are now coming together to form the basis for a mechanism of synaptic transmission. The debate will be further stimulated by the observations of M. Böhler and P. Greengard, reported on page 704 of this issue¹, describing an actin-bundling activity associated with synapsin I. This bundling activity is efficiently expressed only by the dephosphorylated form of synapsin I — on phosphorylation, catalysed by protein kinases activated by nerve depolarization, the activity is lost.

Synapsin I, first discovered by Greengard and colleagues as a substrate for cyclic AMP-dependent protein kinase, seemed to be one of the major substrates for that kinase in the brain. It subsequently became clear that it is also a substrate for calcium/calmodulin-dependent protein kinase and protein kinase C. Most excitingly, though, phosphorylation of synapsin I appeared to occur on depolarization of nerve endings (see ref. 2 for a review). Experiments in which synapsin I in known phosphorylation states was injected into the presynaptic side of the squid giant synapse suggested that synapsin I is involved in the regulation of availability of synaptic vesicles for exocytosis³. Thus dephosphorylated synapsin I might 'restrain' synaptic vesicles in a way that could be overcome by phosphorylation in response to depolarization.

What type of restraint could this be? One possibility is that dephosphorylated synapsin I holds synaptic vesicles in a

cytoskeletal meshwork. If so, phosphorylation would release the vesicle for exocytosis. Böhler and Greengard's observations, then, come in the context of several reports suggesting interactions with cytoskeletal components. Synapsin I apparently interacts with brain spectrin (fodrin)⁴⁻⁶, and the three major classes of cytoplasmic filament: microtubules^{7,8}, neurofilaments^{8,9} and F-actin¹. All these interactions occur *in vitro* with moderate affinity (dissociation constants in the range 0.5–5 μ M). One further activity of dephosphorylated synapsin I is that it enhances the binding of spectrin to F-actin¹⁰, an activity paralleling that of the erythrocyte membrane-bound protein 4.1, with which it has some limited antigenic cross-reaction^{1,5}.

The relationship between synapsin I and actin is perhaps less surprising in view of the recent discovery¹¹ that small regions of the synapsin sequence are 40–50 per cent identical to villin, a calcium-dependent actin severing and bundling protein, and profilin, a protein that sequesters actin monomers. The polymerization reaction demonstrated by Böhler and Greengard is similar to that of profilin under the same conditions. Although Böhler and Greengard have no evidence for an actin-sequestering activity, this is a point that bears re-examination. Equally, it is not clear how synapsin bundles actin — does it have more than one actin binding site per monomer or are monovalent molecules prone to aggregate in the dephosphorylated form to give a polyvalent crosslinking structure? The dissociation constant for actin-synapsin interaction is not related to phosphorylation state, but the stoichiometry of the maximum binding changes from 11 mol actin per 1 mol synapsin in the fully phosphorylated form

to 7 mol actin per 1 mol synapsin in the dephosphorylated form. Such stoichiometries are puzzling but could support aggregation as the means of crosslinking.

Where does this leave us? Because we lack detailed understanding of the cytoskeletal architecture in the presynaptic terminus, it does not yet seem possible to make a coherent model of the relationship (if any) between synaptic vesicles, the cytoskeleton and exocytosis. It should be possible, however, to make a few conjectures. Böhler and Greengard¹ argue that there is little evidence for microtubules and neurofilaments in the presynaptic terminus: indeed, where synapsin has been shown to colocalize with microtubules and neurofilaments it is in parts of the neuron other than the synapse⁸. Conceivably, then, interactions between synapsin and microtubules/neurofilaments may relate more to vesicle transport towards nerve endings than to exocytosis. On the other hand, actin is abundant in presynaptic termini and spectrin is also present. Spectrin seems to be for the most part membrane-bound, so perhaps dephosphorylated synapsin I secures the synaptic vesicle in a complex with spectrin and actin close to the presynaptic plasma membrane. It is also relevant that synapsin *in vitro* can crosslink synaptic vesicles to membranes⁹ and that affinity of synapsin I for synaptic vesicles is reduced by phosphorylation^{9,12}. In this view, nerve depolarization would cause an elevation in free cytosolic calcium activating a calcium/calmodulin-dependent protein kinase and phosphorylating synapsin, thus reducing its interaction with (or at least its ability to crosslink) the synaptic vesicle, spectrin and actin, freeing the vesicle for exocytosis. There are many objections to such a scheme, not the least of which is that synaptic exocytosis occurs extremely rapidly on depolarization. But it seems reasonable to suggest that at the very least synapsin I is involved in positioning synaptic vesicles close to the plasma membrane in a cytoskeletal lattice or meshwork before exocytosis. □

1. Böhler, M. & Greengard, P. *Nature* 326, 704–707 (1987).
2. Browning, M.D., Huganir, R. & Greengard, P. *J. Neurochem.* 45, 11–23 (1985).
3. Llinas, R., McGuinness, T.L., Sugimori, M. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* 82, 3035–3039 (1985).
4. Baines, A.J. & Bennett, V. *Nature* 315, 410–413 (1985).
5. Goodman, S.R. & Zagon, I.S. *Am. J. Physiol.* 250, 347–360 (1986).
6. Kanda, K., Tanaka, T. & Sobue, K. *Biochem. biophys. Res. Commun.* 140, 1051–1053 (1986).
7. Baines, A.J. & Bennett, V. *Nature* 319, 8495–8504 (1986).
8. Goldenring, J.R. *et al.* *J. biol. Chem.* 261, 8495–8504 (1986).
9. Steiner, J., Ling, E. & Bennett, V. *J. biol. Chem.* 262, 905–932 (1987).
10. Krebs, K.E., Zagon, I.S. & Goodman, S.R. *J. Cell Biol.* 103, 2829a (1986).
11. McCaffrey, C.A. & De Gennaro, L.J. *EMBO J.* 5, 3167–3173 (1986).
12. Schiebler, W., Jahn, R., Doucet, J.-P., Rothlein, J. & Greengard, P. *J. biol. Chem.* 261, 8383–8390 (1986).

A.J. Baines is at the Biological Laboratory, University of Kent, Canterbury, Kent CT2 8NJ, UK.

Body temperature and heat and water balance

SIR—In response to J. Paul's¹ suggestion that the body temperatures of homiotherms are all within a few degrees of 36 °C because the specific heat of water is a minimum at 35 °C, Dunitz and Benner² correctly concluded that the specific heat is irrelevant for a body at constant temperature and, on the basis of plotting the viscosity of water and the solubility of benzene in water against temperature, proposed instead that a temperature of 36 °C represented the best compromise. But this conclusion is itself spurious because selection of appropriate scales produces any desired temperature for the intersection of the graphs. We demonstrate here that the optimum body temperature is related not only to the heat balance of animals, as proposed by Calder³, but also to their water balance.

Homiothermy has important consequences for the biochemistry of animals and their ability to harvest food, but the maintenance of a particular temperature will be determined primarily by the balance between the capacities to produce and dissipate heat. The heat generated by metabolism must be dissipated to the environment if body temperature is to remain constant. The body temperature corresponds to the 'working point' of the system, where heat production and loss are in balance. If the rate of heat production depends on body temperature and has a Q_{10} of 2.0, the basal metabolic rate will double for each 10 K rise in temperature. On this assumption, line *a* in Fig. 1 shows how the basal metabolic rate (*M*) per unit area of body surface would vary with body temperature T_b . Rates of metabolism usually exceed the basal level, because of heat production associated with the digestion of food and activity: line

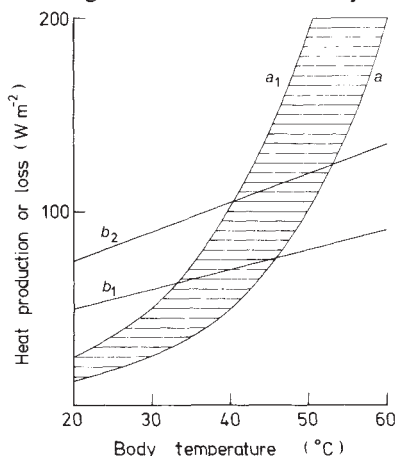


Fig. 1 Heat production and heat loss versus body temperature in the cold (−30 °C). Lines *a* and *a*₁ represent basal metabolic rate *M* and twice *M*, respectively. Lines *b*₁ and *b*₂ represent sensible heat loss for well and moderately insulated animals, respectively.

*a*₁ in Fig. 1 represents twice the basal level, typical of well-fed animals. Heat loss depends on an animal's insulation and its thermal environment. In a cold climate, animals develop enough insulation to ensure that non-evaporative (sensible) heat loss is within their capacity to produce heat. In hot climates, the potential for non-evaporative heat loss is reduced and animals must rely on the evaporation of water to dissipate any excess heat generated by metabolism. We shall now consider how body temperature affects the conservation of energy in the cold, and the conservation of water in a hot environment.

Line *b*₁ in Fig. 1 shows the non-evaporative heat loss plotted against body temperature for a well insulated animal in a cold environment (−30 °C). Clearly, were body temperature below about 35 °C the loss of heat would exceed the normal level of heat production. Conversely, were body temperature above about 47 °C, well insulated animals would need to employ evaporation to dissipate excess metabolic heat. Panting would cause an undesirable increase in metabolic rate, whereas evaporation of sweat would result in condensation within the coat—a problem sometimes encountered by mountaineers and explorers. In the cold, the optimum range of body temperatures therefore corresponds to the part of the heat dissipation line within the hatched area between lines *a* and *a*₁, that is from about 35 °C to 47 °C for a well insulated animal with a minimum evaporative heat loss of 10 W m^{−2}. Line *b*₂ shows the corresponding heat loss from an animal with moderate insulation at −30 °C. For equilibrium, this animal must expend more energy and the minimum body temperature is about 40 °C. The required metabolic rate and body temperature of a poorly insulated animal in a cold environment are unrealistically high, and the food consumption needed to supply the energy expended would be several times that of a well insulated competitor.

In Fig. 2, lines *c*₁ to *c*₃ show sensible heat loss versus body temperature for a homiotherm in a warm environment (30 °C). Line *a* shows *M* plotted against body temperature, and line *a*₂ the part of *M* which must be dissipated as sensible heat for thermal equilibrium, assuming the minimum evaporative loss is 15 per cent of *M*. For warm conditions, we can recognize two cases: first, poorly insulated animals, when the locus for sensible heat loss (line *c*₁) intersects line *a*₂; and second, relatively well insulated animals when line *a*₂ always exceeds sensible heat loss (lines *c*₂ and *c*₃).

For the former, there is a clearly defined 'working point' where the two lines intersect. Here the metabolic cost of homiothermy is the minimum for the

particular value of T_b and the water loss corresponds to the minimum through the skin and in respired air. At lower body temperatures, there would be a small reduction in metabolic rate, but a considerable increase in evaporation to dissipate the difference between metabolic heat production and sensible heat loss. At higher body temperatures, an increase in *M* would be required to match increased sensible heat loss. Perversely, therefore, in warm environments naked animals (line *c*₁), have the lowest metabolic and water needs for thermoregulation. But the optimum body temperature varies with environmental temperature. The optimum T_b is about 34.5 °C at an environmental temperature of 30 °C, but during a cool night the need for heat production would

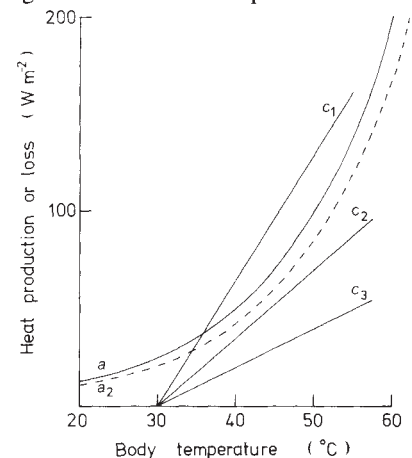


Fig. 2 Heat production and heat loss versus body temperature in a warm environment (30 °C). Line *a*, basal metabolic rate *M*; line *a*₂, *M* minus minimum evaporative heat loss. Lines *c*₁, *c*₂ and *c*₃, sensible heat losses for naked, (poorly insulated) and two examples of relatively well insulated animals, respectively.

be excessive unless the body temperature was allowed to drop. The alternative, which is to maintain a low constant body temperature, would place unacceptable demands on an animal's ability to dissipate heat by evaporation when environmental temperature exceeded T_b , as it often would. Some hair coat is therefore desirable to reduce heat loss at night, even in the tropics, as well as to protect the body from solar radiation during the day.

When line *c* does not intersect line *a*₂, sensible heat loss is always less than metabolic heat production. The evaporative heat loss required for heat balance corresponds to the vertical separation of lines *a* and *c* at the particular body temperature. In this case, there is an optimum value of T_b which gives a minimum water use for thermoregulation. The 'optimum' T_b depends on total insulation, decreasing with increasing insulation. For a few millimetres of hair (line *c*₂) the 'optimum' body temperature is about 40 °C, similar to that of most mammals. Line *c*₃ corresponds to a moderately well insulated animal in the

tropics. For this case, the 'optimum' T_b for water economy is about 32.5 °C, but the lowest corresponding rate of evaporation is several times the minimum possible. Clearly, excessive insulation in the tropics can adversely affect water economy.

We conclude that the observed values of the body temperatures for most homoiotherms, between about 35 °C and 43 °C, are consistent with determination by the requirements of the heat balance. In particular, this range of body temperatures is predicted by considering the need to minimize both energy expenditure in the cold and water loss in warm conditions. High insulation is obviously necessary for homoiothermy in the cold, but a reasonable amount of insulation is also required if body temperature is to be kept constant in a warm environment.

Poorly insulated homoiotherms can be regarded as a different class. In a warm environment, they can minimize both water loss and metabolic rate, but only by permitting a thermolabile body temperature. Therefore, if the dinosaurs had thermoregulated like monotremes they would have been more at risk from sudden climatic changes than relatively well insulated animals, but more favoured in a period of stable climate.

Lastly, in what conditions could homoiotherms evolve on other planets? For an aqueous biota obtaining its energy from the oxidation of carbon compounds and its oxygen from a gaseous atmosphere, similar mechanisms of heat dissipation will occur. The latent heat of vaporization of water changes only slowly over the temperature range of interest, from 2,500 J g⁻¹ at 0 °C to 2,400 J g⁻¹ at 40 °C, so that the potential for evaporative cooling will be similar for all such life forms. Theoretically, the potential for sensible heat loss will depend on atmospheric composition and pressure, but in practice the insulation of fibrous coats is likely to differ little from that of terrestrial animals, because the thermal conductivity of gases changes little with realistic ranges of pressure and atmospheric composition. We conclude that biota employing biochemistry and media similar to those on Earth can only evolve true homoiotherms with a similar range of body temperatures, between about 35 °C and 43 °C. If intelligent life requires a high metabolic rate and homoiothermy, it may therefore be able to evolve only on planets with a relatively narrow range of climates.

ALASTAIR J. MCARTHUR
JEREMY A. CLARK

Department of Physiology
and Environmental Science,
University of Nottingham,
Sutton Bonington,
Loughborough LE12 5RD, UK

A cautionary view of antifertility vaccines

SIR—K.J. Jayaraman's news report¹ of trials of antifertility vaccines comprising gonadotropin hormone subunits and their derivatives calls for some comment. First, it is not quite correct to state that one 'cocktail' of antigens and carriers used by G.P. Talwar for vaccination consists of oLH-hCG-TT (ovine luteinizing hormone-human chorionic gonadotropin-tetanus toxoid) and oLH-hCG coupled to cholera toxoid chain B (CHB), and a second mixture of oLH-TT-CHB and hCG-TT-CHB. In fact, the vaccines either contain the α -subunit of oLH (α -oLH) combined with the β -subunit of hCG (β -hCG) coupled to the two carriers, or β -oLH-TT-CHB and β -hCG-TT-CHB. A third vaccine, which was not mentioned, employs β -hCG-TT alone (for review see ref. 2).

More importantly, doubts arise concerning the hormone specificity, and therefore the safety, of the antifertility vaccines which either already are in phase I clinical trials or are planned to be by the Population Council and the Indian Institute of Immunology. Both groups claim that their carrier-coupled antigens are safe on the basis that α -oLH, β -hCG, β -oLH and α -oLH/ β -hCG are either species-specific (because α -oLH and the α -chain of human glycoprotein hormones are considered immunologically distinct) or, if antibodies against hLH are elicited (because of similarities between β -hLH and β -hCG) they would cause no side effects. The first assumption is based on a concept of Vaitukaitis³ which no longer holds true; only three monoclonal antibody-defined human α -chain epitopes support the assumption⁴, whereas three other epitopes — one of which is only expressed by the non-assembled chain — are not only shared by all α -subunits of human glycoprotein hormones but are also present on α -oLH.

Concerning the immunological specificity of β -oLH and β -hCG chains compared with β -hLH, the situation is even worse. As Jayaraman states, there are numerous reports of immunological cross-reactivities (for example ref. 3). Monoclonal antibody based analyses reveal at least three epitopes shared between the two human β -chains^{4,5}, two of which are also common to oLH. Moreover, monoclonal antibodies against all these shared epitopes are able to inhibit the receptor binding ability of hCG as well as hLH. Additional side effects on the target cell from β -oLH/ β -hCG vaccination might also be caused by antibodies which still have access to the receptor-bound hormone.

One possible way of circumventing the problem of shared epitopes would be to use for vaccination the hCG-specific

carboxy-terminal peptide as done by V.C. Stevens and applied in the WHO supervised Australian trial^{2,7}. If this turns out to be of poor antigenicity, as suggested by Talwar, it would be necessary to investigate further the truly hCG-specific epitopes⁴ and try to isolate them or to construct appropriate mimotopes⁵. All other nonspecific vaccines have to be regarded with scepticism until their safety is clearly proven bearing in mind that they will eventually be applied to millions of people.

PETER BERGER

Institute for General and Experimental
Pathology,
University of Innsbruck Medical School,
Fritz-Pregl-Strasse 3,
A-6020 Innsbruck, Austria

1. Jayaraman, K. *Nature* **323**, 661 (1986).
2. Stevens, V.C. *Immun. Today* **7**, 369–374 (1986).
3. Vaitukaitis, J.L. in *Structure and Function of the Gonadotropins* (ed. McKerns, K.W.) 339–360 (Plenum, New York, 1978).
4. Schwarz, S., Berger, P. & Wick, G. *Endocrinology* **118**, 189–197 (1986).
5. Kofler, R., Berger, P. & Wick, G. *Am. J. reprod. Immunol.* **2**, 212–216 (1982).
6. Geysen, H.M., Barteling, S.J. & Meloan, R.H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 178–182 (1985).
7. Ada, G.L., Basten, A. & Jones, W.R. *Nature* **317**, 288–289 (1985).

Quantitative receptor autoradiography

SIR—Any article which preaches caution in the quantification of autoradiographs is to be welcomed. But if advice is being offered, surely it is important that it should be based on the best information available. We would therefore like to draw attention to some inaccuracies and misleading generalizations that appear in the article by M.D. Hall, A.P. Davenport and C.R. Clark (*Nature* **324**, 493; 1986).

First, the radionuclide ¹²⁵I does not emit β -particles, which by definition are electrons of nuclear origin, but mainly Auger electrons. Contrary to the statement of the authors, most of these (96 per cent) are less energetic than most of the β -particles of tritium (28 per cent at 2.0 keV and 68 per cent at 0.8 keV). It must therefore be assumed that the majority of the higher-energy emissions which sensitize emulsions are X-ray or gamma photons.

Also, in spite of the statement in the article to the contrary, ¹²⁵I-polymer based standards are now commercially available from Amersham International.

The authors further assert that "... the anatomical resolution could be greatly enhanced if the crystal diameter of the emulsion used to make the films was reduced significantly ...". This is grossly misleading. It was shown many years ago by Dr Salpeter and co-workers at Cornell University that even at the electron microscopic level of autoradiography, reducing the halide crystal diameter in the emulsion by a factor of three gives an improvement in resolution (a reduction of

1. Paul, J. *Nature* **323**, 300 (1986).
2. Dunitz, J.D. & Benner, S.A. *Nature* **324**, 418 (1986).
3. Calder, W.A. *Nature* **324**, 418 (1986).

the 'half-distance' of image spread) of no greater than 35 per cent.

J.R.J. BAKER

Ciba-Geigy Pharmaceuticals, Horsham,
West Sussex RH12 4AB, UK

M.A. WILLIAMS

Department of Anatomy and Cell Biology,
University of Sheffield, Sheffield, UK

SIR—Contrary to what Hall *et al.*¹ imply, quantitative autoradiography was developed and used to measure local rates of cerebral blood flow by Kety and co-workers²⁻⁴ well before 1981. This technique has also been applied to determination of local rates of biochemical processes *in vivo*^{5,6}. Quantitative autoradiography, like all techniques, has some pitfalls⁷, but some of the potential errors described by Hall *et al.* are products of their own confusion. Standardization of optical density values against radioactive plastic standards is an accepted method with few problems if it is understood and carried out correctly. Calibration against homogeneously labelled brain tissue sections cut at the prescribed thickness is simple and straightforward. The units to be used are determined by this procedure. They are either d.p.m. or μCi per gram wet weight of tissue. It would be possible to convert values to d.p.m. or μCi per mg tissue protein if concentrations of protein in particular brain regions were known but the reason for the choice of d.p.m. per mm^2 by Hall *et al.* is not clear.

With regard to the problem of relatively short-lived isotopes, an alternative procedure could be devised. Brain sections labelled with the short-lived isotope, such as ^{131}I , could be cross-calibrated against ^{14}C -labelled methylmethacrylate standards. With this approach it is possible to correct for decay mathematically and use a set of standards over and over again.

G. LUCIGNANI

Consiglio Nazionale delle Ricerche,
Centro Studie Fisiologia Del Lavoro
Muscolare,
I-20132, Italy

C.B. SMITH

Laboratory of Cerebral Metabolism,
National Institute of Mental Health,
Bethesda, Maryland 20892, USA

1. Hall, M.D., Davenport, A.P. & Clark, C.R. *Nature* **324**, 493–494 (1986).
2. Landau, W.M., Freygang, W.H., Rowland, L.P., Sokoloff, L. & Kety, S.S. *Trans. Am. neurol. Ass.* **80**, 125–129 (1955).
3. Freygang, W.H. & Sokoloff, L. *Adv. biol. med. Phys.* **6**, 263–279 (1958).
4. Reivich, M., Jehle, J., Sokoloff, L. & Kety, S.S. *J. appl. Physiol.* **27**, 296–300 (1969).
5. Sokoloff, L. *et al. J. Neurochem.* **28**, 897–916 (1977).
6. Smith, C.B., Crane, A.M., Kadekaro, M., Agranoff, B.W. & Sokoloff, L. *J. Neurosci.* **4**, 2489–2496 (1984).
7. Sokoloff, L., Kennedy, C. & Smith, C.B. in *Research Methods in Neurochemistry* Vol. 6 (eds Marks, N. & Rodnight, R.) 79–116 (Plenum, New York, 1985).

HALL AND DAVENPORT REPLY—Baker and Williams are, of course, right to point out that ^{125}I emits electrons and not β -

particles as our original text said*. But, according to Rogers¹, and contrary to the statement of Baker and Williams, electrons derived from ^{125}I have a range of energies, the most abundant of which are: 2.8 keV (27.7 %), 3.6 keV (48.8 %), 22.5 keV (14.2 %), 31.0 keV (6.7 %) and 34.3 keV (1.2 %). In micro- and macroradiography, the relative contribution of these electrons to the blackening of the emulsion is not clear. We can say, however, that under experimental conditions self-absorption of electrons derived from ^{125}I is much less than that of β -emissions from ^3H . In addition, it is generally understood that X rays and γ rays contribute little to the production of an image¹. A ^{125}I polymer-based standard is indeed now commercially available.

With regard to our statement on crystal diameter of the emulsion, Ultrofilm has a reported average diameter of $1.8 \pm 0.3 \mu\text{m}$, whereas Ilford K5 emulsion, used in the coverslip technique, is reported to be $0.2 \mu\text{m}$. This second method affords greater anatomical resolution¹. If a commercial supplier could coat a polymer-based film with, for example, K5 emulsion, the anatomical resolution would be enhanced; we consider 35 % a significant improvement for our application.

In response to Lucignani and Smith, our article was confined to a discussion on receptor autoradiography for brevity, although we were familiar with the use of quantitative autoradiography in related techniques, such as measurement of cerebral blood flow and glucose utilization.

Our reasons for using d.p.m. per mm^2 are quite clear. Three absolute measurements can be made from standard sections, namely optical density, surface area and radioactivity. Using this method it is not necessary to make any assumptions or confuse standard tissue weight per protein with sample protein. We agree that protein measurement within individual brain regions is the ideal situation and a method for performing this has been described².

It is naive to assume that images produced by ^{14}C (β -emission maximum 158 keV; mean 50 keV) can be compared to images produced by ^{131}I , which emits β particles with energy ranges (250 keV, 3 %; 330 keV, 9 %; 610 keV, 87 %; 810 keV, 1 %) ¹.

M.D. HALL

A.P. DAVENPORT

Parke Davis Research Unit,
Addenbrookes Hospital Site,
Hills Road, Cambridge, CB2 2QB, UK

1. Rogers, A.W. *Techniques of Autoradiography* 3rd edn (Elsevier, Amsterdam, 1979).
2. Chiczy-DeMet, A. *et al. Soc. Neurosci. Abst.* **12**, 144 (1986).

*Erratum. Because of errors introduced in the editorial process, the following corrections should be made to the article by M.D. Hall, A.P. Davenport and C.R. Clark (*Nature* **324**, 493–494; 1986): para. 1, line 9—replace ref. 1 by 1–3; para. 5, line 4—replace quantitative by qualitative; para. 6, line 19—replace β -particles by electrons; references—include 5, Clark, C.R. & Hall, M.D. *Trends biochem. Sci.* **11**, 195–199 (1986).

The representation of data in graphs and tables

SIR—Now that Beedle¹ has finally resolved the long-running controversy on the presentation of data in graphs and tables, it is worth going back to understand the origin of the problem.

Each unit used for a physical quantity allows an unequivocal identification of the quantity itself: for example, if the unit is the metre, we are sure that the quantity is a length. If this is good, on the other hand it has revealed a dangerous pitfall as far as our problem goes. On an axis, it is clearly necessary to indicate both the name of the physical quantity and the unit of measure (or multiples or submultiples of either or both). Thus, it is permissible to use

$$\frac{\text{length}}{(\text{metres})}, \frac{\text{length} \times 10^{-3}}{(\text{metres})} \text{ or } \frac{\text{length}}{(\text{metres} \times 10^{-3})}$$

but it is not acceptable to use

$$(\text{metres} \times 10^{-3})$$

to mean one thousandth of the length because this mistakes the unit for the denomination of the quantity.

In fact, if the unit could indicate more than one quantity, we are forced to label the quantity clearly. Thus,

$$\frac{\text{length}}{(\text{metres} \times 10^{-3})}$$

would clearly say that the physical quantity is measured in millimetres. This is the exact opposite of what Luykx² and those who agree with him would indicate. Nobody can prevent them from writing 'c.p.m. $\times 10^{-3}$ ' meaning that these are 'one thousandth of the registered c.p.m.s'. Nevertheless, if we intend to be scientific, let us use the most explicit and simple method of labelling.

MORENO PAOLINI

Istituto di Farmacologia,
Università degli Studi,
Via Irnerio 48,
40126 Bologna, Italy

GIOVANNI CERCIGNANI

CARLO BAUER

Dipartimento di Fisiologia e Biochimica,
Università degli Studi,
Via S. Maria 55, 56100 Pisa, Italy

1. Beedle, A. S. *Nature* **325**, 305 (1987).
2. Luykx, P. *Nature* **315**, 462 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

Relics of Hutchins and Fermi linger on

Chicago, a city boasting more than its share of Nobel laureates, is a tempting prospect for would-be members of the 'brain-drain'.

Chicago (March)

Is the glut of academics about to end? This is the surprising indicator of the academic labour market as seen from the University of Chicago, where the provost, Norman Bradburn, says that it has become increasingly difficult to fill faculty positions with people from the United States, which is why "we're looking closely to Oxford and Cambridge".

The explanation is only partly demographic. The expansion of the US university system after the Second World War, driven by the rights of demobilized servicemen to get themselves a college education, began earlier than in Britain. This means that large numbers of academics will soon be reaching retirement age. So the need for academics, at least at places such as Chicago, remains high. What more natural than to extend recruitment networks to other English-speaking countries where would-be academics are in plentiful supply?

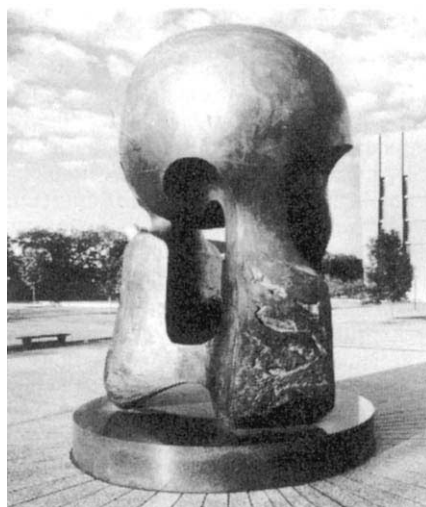
To make seduction simpler, the University of Chicago reckons now to have lived through and recovered from its troubles, in the 1950s, when it became the surprised victim of the urban blight and social unrest in which it was physically embedded. Then, according to one member of the faculty, the half-life of a newly appointed faculty member was about 150 days. Another explains that people were unwilling to travel away to meetings if their wives were required to stay behind. But now urban blight has been beaten back by the university's deliberate involvement in urban development, so that 90 per cent of the faculty are now able to walk to work.

The campus is even better known as the place at which Fermi built a nuclear reactor (the first) in a squash court without telling the university president what he was about. The spot is now marked by a Henry Moore sculpture whose symbolism remains a matter for academic conjecture. A more direct legacy is the Enrico Fermi Institute, physically an evidently 1940s building but intellectually a kind of living museum of Fermi's enthusiasms — cosmic rays, isotope chemistry and particle physics, for example.

But Fermi's post-war cyclotron has been dismantled and the hangar-like space that housed it cannibalized for the building of particle detectors for use at Fermilab (the National Accelerator Laboratory at Batavia, the other side of the city); in one corner is a shed (to keep out dust) housing a detector of primary cosmic rays returned from one of the

earlier shuttle flights and waiting for another chance to fly.

Chicago is also the private university (outside the Ivy League) at which, between 1929 and 1951, Robert M. Hutchins made his name as a distinctive university president, most memorably as an advocate and practitioner of a broad liberal arts education as a foundation of all curricula. Hutchins' legacy is not as easily identified as Fermi's, but it remains. Undergraduates in their first two years must follow courses in each of the physical, biological and social sciences and in the humanities, learn a foreign language and some mathematics (but the calculus is about to give



Conjectural symbolism.

way to 'quantitative reasoning') as well as a course in civilization. But the reading of Hutchins's Great Books (with Adler) has been abandoned as unteachable. In some respects, the cry is "Long Live Hutchins! Hutchins is dead!"

But Hutchins' interdisciplinarity conspicuously remains. Cosmochemist Edward Anders, formally of the Chemistry Department, might elsewhere be labelled a physicist or even a geologist. Astronomy and astrophysics seem to be part of a seamless web including physics. Albert V. Crewe, building an electron microscope that he reckons will have a resolution of 0.5 ångström (thanks to the clever design of a sextupole magnetic lens) has biologist Serge N. Vinogradov writing programs for the computer equipment that is a large part of the \$1 million grant from IBM with which the microscope is being built. The principle seems to be that it is more prudent that a department should acquire critical masses of able people than to attempt to cover the whole of what would

conventionally be called its field.

Hutchins' liberalism also shows in the university's declared policy of what it calls "needs-blind" admissions policies. The boast is that nobody considers how much a student will contribute towards tuition costs when considering applications. One consequence is that Chicago's spending out of its own resources on scholarships and fellowships has increased to more than a fifth of the total. Another is that the cost of tuition has risen steeply, from \$9,600 last academic year to \$12,000 next.

There are other clouds on the horizon. One product of the troubles of the 1950s and 1960s was the analysis that the university had grown too quickly; the response, according to provost Bradburn, was to "shrink a little", but by the 1980s the university was alarmed that the numbers of graduate students had fallen, "precipitously" in some departments, physics for example. In 1982, a Commission on Graduate Education decided that the explanation was not demographic and recommended various remedies, some educational and some administrative.

On balance, the changes appear to have worked. Graduate recruitment has improved in the departments that formerly lagged behind. After the earlier shrinkage, it is also planned to increase the intake from high schools to reach a total of 3,400 undergraduates (compared with close on 3,000 now) without a corresponding increase of the faculty strength (just over 1,100). The result should be a total student body of some 10,000, divided roughly three ways between undergraduates, graduate students and students in the professional schools.

Surprisingly, given its research record (from Michelson, Arthur Compton and Mulliken in the old days to Chandrasekhar still), the university is also worried about research income, which amounted to \$62.4 million from federal agencies in 1985–86 (together with \$28.1 million in indirect or overhead costs). This is a smaller proportion of the total educational budget (just over \$250 million) than it might be, but Bradburn is especially concerned that Chicago, which does not have an engineering school, will be the loser at a time when federal agencies are increasingly concerned with engineering research. Perhaps a little late in the day, the university has just started a computer science department. It is far from being the only excellent institution in the United States now suddenly more anxious about its future than for 30 years. **John Maddox**

Life on the high wire

Andrew Watson

The Evolution of the Biosphere. By M. I. Budyko. *Reidel*:1986. Pp.423. £59.75, \$94.50.

I HAVE never met Mikhail Budyko, but the measured and scholarly style of his book conjures up the image of a distinguished academic of the old school. His delivery may be a little didactic, but the writing is clear and to the point. He manages to cover a sizeable subject area, including many topics about which little is known and too much is speculation, without once losing the aura of rational academic authority. Budyko is most respected for his contributions to climatology, and it is in this field that he is at his most informative. But he is almost equally at home when discussing palaeontology, geology, evolution or ecology. The main strength of the book is in Budyko's ability to slip easily between these subjects and to show how intimately they must be integrated for the proper study of the biosphere. All readers, except the few who have mastered these subjects as well as Budyko, will find plenty to interest them here.

Ultimately, however, the author's failure to establish his interpretation of the evolution of the biosphere as the correct one also stems from the attempt to cover a huge area from too-authoritative a standpoint. Budyko would like us to believe we are reading a textbook, where the subject has been tamed and is no longer open to serious doubt. But some of the most important conclusions are based on oversimplified models which may be giving thoroughly erroneous results. Simple models can be very useful. But Budyko often fails to let his readers know that they are being given only part of the story.

The book is divided into three sections. The first five chapters describe various aspects of the present state of the biosphere, which is defined as "the zone in which modern organisms live" — roughly the lower atmosphere, hydrosphere, and Earth's surface and soil. Starting from the point of view that temperature and water supply are the 'master variables', the author uses climatology theory to explain the distribution of vegetation on Earth. Most of this material is uncontroversial, so the detailed textbook approach is entirely appropriate. Some of the detail is in fact a little excessive: there are eight pages on calculating the zone of the Soviet Union in which a naked man will be comfortable outdoors in summer. This might be very useful if the airline loses your suitcase on a trip to Moscow, but it is only peripherally relevant to the main subject.

The central part of the book deals with

the history of the biosphere. Here Budyko abruptly leaves the well-founded subject of modern climatology to deal with problems where the data are decidedly sparse. The 'master variable' in this discussion is the chemical composition of the atmosphere, particularly the concentrations of oxygen and carbon dioxide. Precambrian time, which accounts for the first five-sixths of the period which life has been present on Earth, is given short shrift. This is understandable, given the almost total absence of hard facts for this period. Budyko supports the view that free oxygen first appeared over 2×10^9 years ago but for some reason remained at very low levels until the late Precambrian, while carbon dioxide probably decreased steadily.

For the Phanerozoic, the last 6×10^8 years for which a far richer fossil record exists, Budyko constructs histories for the oxygen and carbon dioxide content of the atmosphere which he then proceeds to relate to the diversity of life. The approach appears at first to be quite successful, but on closer examination the success is more apparent than real. The starting point for the oxygen calculation is the amount of organic carbon buried in sediments. This is stoichiometrically equal to the net oxygen production from the biota (that is, photosynthesis minus respiration). For the oxygen sink, Budyko simply takes the rate of removal by

weathering to be proportional to the mass in the atmosphere. This is a very shaky assumption given that the oxidative weathering of most minerals goes to completion at very low oxygen pressures. It also ignores changes in rates of physical erosion. No mention is made of the variations of the extent of sulphide and sulphate reservoirs, which can be deduced from isotope data and which have a considerable effect on the calculation.

From these assumptions, Budyko concludes that oxygen concentrations are high at times of high organic carbon sedimentation. He then notes that such periods (such as the Devonian and Carboniferous) seem to be times of high diversity among organisms, and he asserts that this must be because the atmosphere was so rich in oxygen. But a more straightforward argument would be that more organic carbon is being trapped in sediments *because* there is an abundant and diverse biota.

For carbon dioxide, the modelling is even simpler: the concentration in the atmosphere is taken to be proportional to the total carbon content of sedimentary rocks per unit time. There is no real physical justification for this assumption, the result of which is high atmospheric carbon dioxide and hence, through the 'greenhouse' effect, higher temperatures at times of enhanced carbonate sedimentation. Now, it is true that carbonate sediments tend to be laid down during warmer climatic periods; but *why* this should be the case is simply not answered by Budyko's model.

For both oxygen and carbon dioxide, Budyko seems to be telling only half of the story: he discusses how life on Earth may respond to changing atmospheric



*Flying deception — an Edwardian postcard composed of a photograph of Stonehenge with a Bristol Boxkite biplane and monoplane superimposed. The picture is reproduced from the new paperback reprint of Christopher Chippindale's *Stonehenge Complete*, published by Thames & Hudson. Price is £9.95.*

conditions but largely ignores the effect that the organisms themselves may have on the atmosphere. He is describing the *evolution* of climate and life, but not their *co-evolution* which finds its most radical expression in Lovelock's Gaia hypothesis. Perhaps as a result of this different perspective, Budyko's analysis leads him to believe that the biosphere is only marginally stable: that it could be destroyed by relatively small changes in, say, solar insolation.

The question here is, why has life on Earth survived for so long? Is it simply that a series of lucky coincidences has kept conditions within the narrow limits of temperature and atmospheric composition which enable life to continue? Or is the system inherently so stable that no chance was involved? Or, as Lovelock argues, is it life itself which has kept the climate equitable? Budyko is one of the 'lucky chance' school. In particular, in the depths of the most extreme glaciations, the planet may have been close to the 'white Earth' catastrophe: the increase in albedo from extension of the polar ice sheets reflects more radiation away from the planet, resulting in a powerful positive feedback which further intensifies the cooling. If the glaciation is intense enough the whole process can in theory run away until the ice-sheets cover the Earth. At the other extreme, it might not take too great an increase in the absorbed radiation to trigger a 'runaway greenhouse' in which the oceans would boil into the atmosphere. Between these hells of eternal winter and superheated steam, life on Earth has apparently so far kept its balance by nothing but the darndest good luck. It is a breathtaking high-wire act, all the more spectacular because at frequent intervals (geologically speaking) asteroids up to tens of kilometres in size collide with the planet, causing massive extinctions.

In the last section, Budyko deals with the relationship between man and the biosphere. He seems almost to welcome the onset of the 'greenhouse' caused by anthropogenic emissions of carbon dioxide, as a way of increasing the mean temperature and thus reducing the adverse effects of future glaciations. Most scientists would undoubtedly disagree with him on this issue, but with his final appeal I for one am in complete accord: a global nuclear war would result in a trauma to the biosphere of unpredictable magnitude but probably comparable to the mass extinctions of the past. Compared to the effects of such a conflict, other present-day threats to the biosphere are minor. Nuclear disarmament is the most significant contribution to the continued health of our planet that modern man can make. □

Andrew Watson is a Principal Scientific Officer at the Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK.

Works of another Darwin

Willem Hackmann

Horace Darwin's Shop: A History of the Cambridge Scientific Instrument Company 1878–1968. By M. J. G. Cattermole and A. F. Wolfe. *Adam Hilger*: 1987. Pp. 285. £35, \$77.

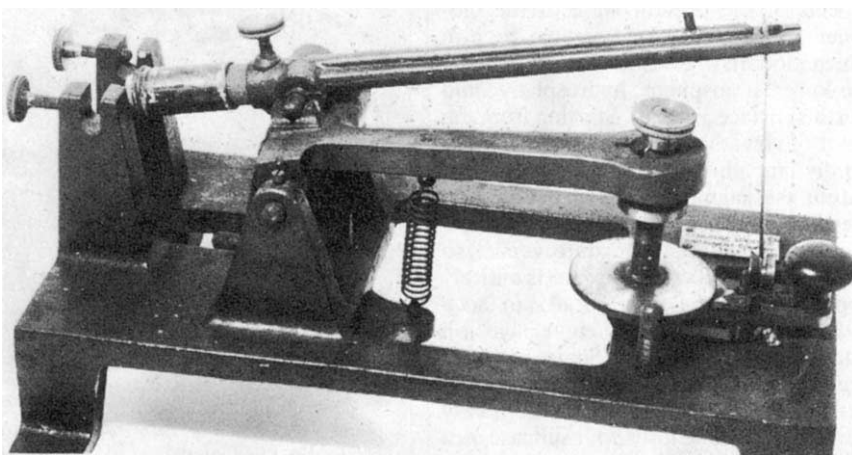
THE Great Exhibition of 1851, intended to show the triumph of Britain's industrial might, instead provided evidence that in certain respects the nation was being eclipsed by the technical capabilities of her European rivals. The country's leaders were spurred to improve technical and scientific education at schools and universities, the effects of which began to be felt in the 1870s. At Cambridge University the rapidly developing Departments of Experimental Physics and of Physiology provided the stimulus for the establishment of a local instrument-making firm. This was not a unique development: James Watt, a leading figure in the Industrial Revolution, had begun as instrument-maker to Glasgow University; in the 1820s Francis Watkins, a partner in Watkins & Hill, famous for their electromagnetic apparatus, became the 'curator of philosophical instruments' for London University; and later, the instrument-maker James White formed a fruitful partnership with Lord Kelvin.

This book recounts the humble beginnings of instrument-making in Cambridge, for which the university authorities, with minimal experience in scientific research, can take little credit. As is so often the case with new ventures, it was begun by a few enthusiasts: James Stuart and Michael Foster, respectively the Professor of Mechanisms and Trinity Praelector in Physiology, Albert Dew-Smith and Horace Darwin. Foster, hampered by a

chronic shortage of funds for purchasing expensive physiological apparatus from the continent, was helped out financially on many occasions by his friend, the wealthy Dew-Smith. Stuart established a departmental workshop (which was to grow into the Engineering Laboratories of the university) for teaching students and manufacturing badly needed instruments. These were sold to the university to defray the costs of wages for the two mechanics he employed as instructors. One of these was Robert Fulcher, who was given permission to make apparatus commercially, for which Dew-Smith provided the financial backing. When this arrangement proved unsatisfactory, Fulcher and Dew-Smith left the Stuart workshop and set up on their own in a hayloft. In 1881 this partnership broke up and was replaced by one consisting of Dew-Smith and Horace Darwin, trading under the name of the Cambridge Scientific Instrument Company, but always referred to by Darwin's family as 'Horace's Shop'.

Horace Darwin was born during the year of the Great Exhibition into a remarkable family. His father was Charles Darwin, his great-grandfather Erasmus Darwin and Francis Galton was his second cousin. He obviously had instrument-making in his blood, for another of his great-grandfathers was Josiah Wedgwood, the pottery magnate and inventor of the pyrometer.

The book is divided into two parts. Part 1 is a fascinating account of the development of the company, which drew much of its inspiration for new instruments from the research taking place in the university. This symbiotic relationship between commercial instrument-making and university research (the present day 'science park') has become commonplace, but in the early decades of this century was a novel departure. The successes and failures of the Cambridge Scientific Instrument Company, and the many recent mergers and take-overs (ten between 1970



'Darwin's Rocker' — a design of microtome first introduced by the Cambridge Scientific Instrument Co. in 1885, and subsequently used by generations of microscopists.

and 1985), mirror what has happened to British technology and industry in general.

Part 2 describes in greater detail some of the instruments developed by the company, from the automatic microtome ('Darwin's Rocker' of 1885) to the Cambridge cardiograph and Wilson's cloud chamber. Some of these devices have had a remarkably long life, and in the 1950s the firm became rather conservative in its designs, resisting for instance the introduction of electronics — as one of the old workmen put it: "This isn't instrument

work — it's ironmongery!"

Horace Darwin's Shop is of interest both as a company history and for its accounts of developments in scientific instruments. It deserves a wider readership than the 'buffs' interested in katharometers, photobustoscopes (Darwin's name for a camera system for recording explosions), photonephometers and auxanometers. □

Willem Hackmann is Assistant Curator at the Museum of the History of Science, University of Oxford, Broad Street, Oxford OX1 3AZ, UK.

Beastly behaviour

Tim Clutton-Brock

Ecological Aspects of Social Evolution: Birds and Mammals. Edited by Daniel I. Rubenstein and Richard W. Wrangham. Princeton University Press: 1986. Pp.351. Hbk \$65, £43.40; pbk \$23.50, £15.70.

AS THE diversity of animal social behaviour became apparent in the 1950s and 1960s, interest was initially concentrated upon the reasons for differences between species. Research in this area crystallized when, in 1964, John Crook demonstrated that variation in sociality and breeding behaviour among weaver birds was consistently related to differences in habitat type and diet. Subsequent analyses revealed similar correlations between social behaviour and ecology in many other groups of vertebrates.

Crook's approach to understanding the adaptive significance of gross differences in social behaviour by investigating associations between species differences in ecology and social behaviour was termed socioecology. In the mid-1970s, it was largely (though not totally) eclipsed by the birth of sociobiology with its associated emphasis on the reproductive strategies of individuals. This generated a new wave of field work which focused on the competitive and cooperative interactions of individuals within groups, revealing new orders of complexity in social behaviour.

Ecological Aspects of Social Evolution marries the sociobiological and socioecological approaches in an attempt to understand the adaptive significance of species differences in social behaviour. The first section consists of eight chapters describing the behaviour of primarily monogamous or polyandrous species or groups, while the second includes nine chapters on polygynous animals and one on human breeding systems. Most of the authors use detailed studies of single species as a springboard for explanations of differences in social behaviour between related species. The contributions are of a uniformly high standard, and together they provide an illuminating view of

recent research on the subject.

Two themes recur throughout the book. The first is the extent to which social and reproductive behaviour are influenced by demography. For example, the retention of juveniles or subadults as helpers within the parental territory is usually associated with low adult mortality rates and a corresponding increase in the costs of dispersal in saturated environments. An increase in adult mortality (either within or across species) is commonly associated with an increase in dispersal and a reduction in group size and the number of

the evolution of social behaviour are repeatedly forced to invoke costs and benefits which can be guessed at but seldom measured directly, and, instead, we are left to rely on indirect evidence. While this can be compelling, it rarely allows us to compare the strengths of different selection pressures and it is often difficult to distinguish credible arguments from castles in the air. For example, in primates there is not much direct evidence to support the oldest of all explanations of sociality: that grouping reduces predation rate. This lack of evidence has been used to imply that social advantages of grouping predominate. But consider the probability of a field worker demonstrating an association between group size and predation rate in a population of animals where most individuals live for ten years and 50 per cent are eventually killed by predators. It is rare for a researcher to average more than 200 hours of observation a month and be able to monitor more than 20 animals at a time. Even if we assume that predators only attack by day, a field worker will only witness *one animal being killed every year*. Under such conditions, it is virtually impossible to produce

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

In at the kill — but lionesses may be social animals for reasons other than hunting.

helpers. Such findings emphasize that many apparently species-specific aspects of social behaviour may be flexible responses which can vary with local environmental conditions.

The second theme is the extent to which diverse aspects of social behaviour may be adapted to the behaviour of conspecifics rather than to food distribution or predation, the two ecological parameters most commonly considered by socioecologists. Among felids, for example, lions are the only species in which females are social. It has often been suggested that grouping in lionesses represents an adaptation to maximizing hunting success. However, Packer's results, summarized in the penultimate chapter, show that meat intake declines with increasing group size. The main benefit of sociality, he suggests, is that by increasing pride size lionesses increase the tenure of resident males and reduce the frequency of infanticide.

The book, however, allows little room for self-congratulation amongst those of us working in this field. Arguments about

direct evidence that the frequency of predation declines with increasing group size, and the absence of such evidence cannot be used to suggest that predation is unimportant as a cause of sociality.

As Wrangham and Rubenstein point out in the final chapter, long-term studies of the reproductive costs and benefits of different social strategies will be needed to provide firm evidence of adaptive explanations of variation in social behaviour among birds and mammals. In particular, parallel studies of the fitness pay-offs of different strategies in related species with contrasting social systems may provide revealing insights into the adaptive significance of species differences, while, at a different level, quantitative interspecific comparisons will continue to play an important role in allowing us to test hypotheses about gross differences in social behaviour. □

Tim Clutton-Brock is in the Large Animal Research Group, Department of Zoology, University of Cambridge, 34A Storey's Way, Cambridge CB3 0DT, UK.

Growing up in the permafrost

Robert Foley

Reindeer Moon. By Elizabeth Marshall Thomas. Houghton Mifflin: 1987. Pp. 338. \$17.95.

WHATEVER you do, don't let your daughter marry a mammoth hunter. They beat their wives, they think about nothing but their work (hunting) and they tell boring stories about their relatives. Worse, they make their wives wear baggy trousers, not only a travesty of fashion for a self-respecting woman of the Fire River, but also impractical if you have to climb a tree to escape a lion.

Of course, the opportunities for marrying mammoth hunters are limited, for they lived 20,000 years ago and in all probability exist only in the imagination of Ms Thomas, the author of *Reindeer Moon*. In the usual sort of *Nature* review this last comment would be damning criticism, but in this case it is not.

Reindeer Moon is a novel. It tells the story of Yannan, a young girl growing up in the permafrost. Orphaned, starved and married, she survives all these hazards of the Eurasian steppes during the last ice age; survives until page 27, that is. From then on there are really two tales running in parallel — one, the story of Yannan as a young woman facing the tribulations of a harsh natural and social environment; the other, Yannan transformed into a spirit after death, looking down on the incessant struggle for food continued by the survivors.

Her life was certainly hard as well as short, and this book is by no means a representation of prehistoric man as the noble savage that one might have expected from the author of *The Harmless People*. Any harmony that human beings had with nature derived solely from the fragile hold on life that they must have had in the bleak and endless steppes of the last glaciation. The story of Yannan's life is exhausting — ever-present hunger, satisfied only momentarily by hunting, a relentless need for firewood to ward off the biting cold, with predators, septicaemia and death in childbirth as more intermittent risks in case these others relent. The price of survival is constant social interaction — communal living to maintain warmth, economic interdependence, and marriages and alliances forced by the need for access to resources. The association between the people of the Fire River and the boorish mammoth hunters of the adjacent valley is one such alliance, bringing with it security of food at the cost of social and emotional conflict. Biting arctic winds and oppressive social claus-

trophobia alternate as the pervasive themes of the book.

To some extent Ms Thomas has written a Pleistocene soap opera, an up-market *Clan of the Cave Bear*. But her book is more than that. She manages to evoke the beauty of these arctic environments without losing sight (or, more particularly, the smell) of their hostility. Her research has been painstaking, incorporating a detailed knowledge of the habits and habitats of the plants and animals of Eurasian steppe with some imaginative anthropological reconstruction. The force of accuracy is weakened by the fact that the spirit world after death is presented with the same degree of realism as the lost prehistoric world, but the book imparts vividly the atmosphere of living at that time. *Reindeer Moon* started as a piece of scientific research into prehistoric hunter-gatherer ecology, and this may account for

its seamless continuity between fact and fiction.

The Reindeer Moon of the title refers to the name of one of the months, and the story passes through many moons, each reflecting an annual event — the ice-breaking moon, the yellowleaf moon, the hunger moon and so on. At first this is confusing, but it is soon infectious and one realizes how remote the names of our months are from the realities of everyday life. It is time, perhaps, to revitalize the academic calendar to reflect the trials of contemporary life. So how about the Month of Grant Proposals Due, the Month of Disappointment, and that hardy British perennial, the Month of UGC Cuts? □

Robert Foley is a Lecturer in the Department of Physical Anthropology, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

Repair in the dark

Tomas Lindahl

Molecular Biology of DNA Repair. Edited by A. Collins, R.T. Johnson and J.M. Boyle. *The Company of Biologists, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK: 1987. Pp. 353. £35, \$70.*

IN THE celebrated concluding statement to their first 1953 paper on DNA, Watson and Crick commented that "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material". This remark and the subsequent discussions of DNA replication may be contrasted with an afterthought by Crick in a review 21 years later: "We totally missed the possible roles of enzymes in repair".

DNA repair has continued to escape the attention of most molecular biologists, as witnessed by the cursory and erroneous coverage of the subject in the textbook *Molecular Cell Biology* by Darnell, Lodish, and Baltimore, published last year by W.H. Freeman. This state of affairs is unfortunate and surprising. Cells suffer so many forms of spontaneous and induced damage to their genomes that a multitude of different pathways and mechanisms of repair have evolved, and these processes are instrumental in preventing unacceptable levels of mutagenesis and carcinogenesis. Indeed, a larger part of the cellular enzymatic machinery for nucleic acid metabolism seems to have been devoted to working out the many small problems of DNA repair than to solving the single towering question of DNA replication.

Although occasional highlights of re-

pair studies may appear in the better-known publications, much of the solid groundwork in this field is to be found in cancer journals or those specializing in mutation and radiation research or genetic toxicology. An authoritative overview of the molecular biology of DNA repair was long needed, and arrived in the form of Errol Friedberg's excellent *DNA Repair* (published by W.H. Freeman in 1985). This is the book against which all others on the subject have to be measured.

The volume under review consists of a series of articles and reviews representing contributions to a conference held in Manchester in 1986. Compared with Friedberg, the coverage is patchy — some aspects of the field are usefully brought up to date, but on others, such as photoreactivation where the previous book is already out of date, there is no information at all. Most of the contributions come from European laboratories and many are of uneven quality.

There are, however, a thoughtful concluding review on mutagenesis and cell killing in mammalian cells by methylating agents, and informative articles on inducible DNA repair systems and excision repair in *Drosophila* and human cells. Also, the editors (who somewhat pretentiously refer to themselves as the DNA Repair Information Network) have compiled a useful and extensively referenced catalogue of the many different DNA repair mutants that have been described in higher eukaryotes.

The book is elegantly produced. Taken as a symposium volume it is somewhat fragmentary in content, but above average in quality. □

Tomas Lindahl is Head of the Clare Hall Laboratories, Imperial Cancer Research Fund, South Mimms, Hertfordshire EN6 3LD, UK.

Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate

Robert J. Charlson*, James E. Lovelock†, Meinrat O. Andreae‡ & Stephen G. Warren*

* Department of Atmospheric Sciences AK-40, University of Washington, Seattle, Washington 98195, USA

† Coombe Mill Experimental Station, Launceston, Cornwall PL15 9RY, UK

‡ Department of Oceanography, Florida State University, Tallahassee, Florida 32306, USA

The major source of cloud-condensation nuclei (CCN) over the oceans appears to be dimethylsulphide, which is produced by planktonic algae in sea water and oxidizes in the atmosphere to form a sulphate aerosol. Because the reflectance (albedo) of clouds (and thus the Earth's radiation budget) is sensitive to CCN density, biological regulation of the climate is possible through the effects of temperature and sunlight on phytoplankton population and dimethylsulphide production. To counteract the warming due to doubling of atmospheric CO₂, an approximate doubling of CCN would be needed.

CLIMATIC influences of the biota are usually thought of in connection with biological release and uptake of CO₂ and CH₄ and the effect of these gases on the infrared radiative properties of the atmosphere¹. However, the atmospheric aerosol also participates in the radiation balance, and Shaw² has proposed that the aerosol produced by the atmospheric oxidation of sulphur gases from the biota may also affect climate. So far the physical and biological aspects of this intriguing hypothesis have not been quantified, but three recent discoveries may make this possible for the remote marine atmosphere. (1) Most species of phytoplankton, ubiquitous in the oceans, excrete dimethylsulphide (DMS) which escapes to the air where it reacts to form a sulphate and methane sulphonate (MSA) aerosol. (2) This non-sea-salt sulphate (NSS—SO₄²⁻) aerosol is found everywhere in the marine atmospheric boundary layer. (3) Aerosol particles which act as cloud-condensation nuclei (CCN) in the marine atmosphere are principally, perhaps almost exclusively, these same NSS—SO₄²⁻ particles.

In this paper we show that emission of DMS from phytoplankton is sufficient to justify its consideration as the gaseous precursor of CCN in the remote and unpolluted marine atmosphere. We re-examine the physical role of the sulphate aerosol in atmospheric radiative transfer, particularly in clouds, which are responsible for most of the Earth's albedo, and estimate the sensitivity of the Earth's temperature to changes in the abundance of CCN. Finally we examine the geophysics³ of the system comprising phytoplankton, DMS, CCN and clouds, as a putative planetary thermostat.

Production of sulphur-containing gases

Assuming that sulphuric acid and sulphate aerosols are indeed the only significant contributors to CCN over the oceans (as will be concluded below), we have to consider what processes are responsible for the production of the volatile sulphur compounds that are the precursors of sulphuric acid in the atmosphere. In this discussion, we shall ignore the perturbations of the atmospheric sulphur cycle by manmade fluxes of SO₂ (mostly from burning of coal and oil) and shall consider only the natural fluxes, which currently represent about 50% of the total flux of gaseous sulphur to the atmosphere, and which still dominate the atmospheric sulphur cycle in the Southern Hemisphere⁴⁻⁶.

The only significant non-biological natural flux is the emission of SO₂ and H₂S by volcanoes and fumaroles. This process releases of the order of 0.4 Tmol yr⁻¹, about 10–20% of the total natural flux of gaseous sulphur to the atmosphere^{4,5}. The emission of sulphur by volcanoes is highly variable in space and time. Consequently, the production of sulphate aerosol by the oxidation of volcanic sulphur during the quiescent stage of

a volcano is of regional importance only. Large eruptions, on the other hand, which emit enough gaseous sulphur compounds to influence wider areas, are relatively rare events. For this discussion, we shall assume that the contribution of CCN from volcanic sulphur to the global atmosphere is proportional to its contribution to the total sulphur flux, that is, 10–20% of the natural component at present.

Volatile sulphur compounds are emitted both by terrestrial and marine biota. The marine emissions are almost exclusively in the form of DMS, whereas the emissions from land are in a variety of chemical species, including H₂S, DMS, methanethiol, CS₂, COS and others. This difference is related to the biological processes which are responsible for the production of the sulphur volatiles. Most sulphur at the Earth's surface is present as sulphate, which is the thermodynamically stable form of sulphur in the presence of oxygen. Sulphate is reduced by organisms through two mechanisms, 'assimilatory' and 'dissimilatory' sulphate reduction. The dissimilatory pathway is restricted to sulphate-reducing bacteria in anaerobic environments; due to physical and microbial restrictions only a small fraction of the H₂S produced by this process can escape to the atmosphere⁷. The products of the assimilatory pathway are a variety of organosulphur compounds, the largest fraction being the amino acids cysteine and methionine, which are incorporated into proteins.

On the continents, the breakdown of organosulphur compounds during fermentative decomposition of organic matter is probably the most important mechanism for the release of sulphur volatiles. Due to the difficulty in obtaining representative data from land biomes, the magnitude of this flux is still poorly known. Estimates range from about 0.15 to 1.5 Tmol yr⁻¹; the best estimate appears to be around 0.25 Tmol yr⁻¹ for the total land surface, or about 2 mmol m⁻² yr⁻¹, which turns out to be a slightly smaller flux per unit area than is emitted from the oceans (about 3 ± 1.5 mmol m⁻² yr⁻¹). Preliminary data from the tropical regions suggest that there may be a net transport of biogenic sulphur from the marine to the continental atmosphere, consistent with a smaller emission flux per unit area on the continents⁸. Coastal wetlands play a surprisingly small role in the global sulphur cycle; at present we estimate their contribution to be about 2% of the total gaseous emissions⁵. This is because although the emission per unit area may be relatively high, the total area of coastal wetlands is quite small.

In the marine environment, most volatile sulphur is emitted in the form of DMS which is excreted by living planktonic algae. In contrast to microbial decomposition which produces a complex mixture of volatile sulphur species, algal metabolism yields DMS as the only volatile sulphur species. The biological func-

tion of the production of DMS is still unclear. The substance from which DMS originates, dimethylsulphonium propionate (DMSP), is important in osmoregulation in a number of phytoplankton types⁹ and also participates in the biochemical cycle of methionine. DMSP is also excreted by algae, and its breakdown in sea water may release additional amounts of DMS.

Even though DMS is excreted by phytoplankton, its concentration in surface sea water, and consequently its emission rate to the atmosphere, is only weakly correlated with the usual measures of phytoplankton activity, for example, chlorophyll concentration or ¹⁴C-uptake rate. In fact, the calculated flux of DMS from the tropical oceans, which have a low primary productivity, is essentially the same per unit area (2.2 mmol m⁻² yr⁻¹) as from the much more productive temperate oceans (2.4 mmol m⁻² yr⁻¹)⁵. Recent evaluations of seasonal effects^{10,11} suggest that the annually averaged flux from temperate regions may even be considerably lower than the value given here. Even the highly productive coastal and upwelling regions do not support much higher DMS fluxes (5.7 and 2.8 mmol m⁻² yr⁻¹), respectively, based on a large body of data from several research groups⁵). It appears that, independent of the rate of primary production, the warmest, most saline, and most intensely illuminated regions of the oceans have the highest rate of DMS emission to the atmosphere. This is a key fact to keep in mind when discussing possible climatic feedback mechanisms.

Two reasons can be given for this unexpected behaviour. First, the concentration of DMS in surface sea water depends not only on the rate of DMS production, but also on its rate of removal. The two dominant removal processes are ventilation to the atmosphere and the photochemical and microbial breakdown of DMS in the water column^{12,13}. The rate of microbial removal of DMS will increase both with the density of bacteria which are able to metabolize DMS and, in a nonlinear fashion, with the concentration of DMS in sea water. The density of bacterioplankton is a function of the densities of phytoplankton and of grazing organisms. Due to these diverse and nonlinear relationships between the variables which regulate DMS production and removal, we cannot expect to find a simple, linear relationship between phytoplankton density and DMS concentration. Furthermore, the production rate of DMS shows variations over three orders of magnitude among different phytoplankton species (ref. 14 and M.O.A., unpublished data). It appears that some algal groups, such as the coccolithophorids, which are most abundant in tropical, oligotrophic waters, have the highest rate of DMS excretion per unit biomass. We conclude, therefore, that the global input of DMS into the atmosphere is much less sensitive to changes in total primary production than to variations in phytoplankton speciation. In other words, we can accommodate rather large swings in marine productivity without changing the climatic effects produced by the sulphur cycle, and on the other hand, we could essentially eliminate the marine input of sulphur to the atmosphere without significantly changing the total primary productivity, for example by replacing all coccolithophorids with diatoms. The concentration of SO₄²⁻ in sea water is so large (~29 mmol kg⁻¹) that it does not limit the emission of DMS.

We can now evaluate the role of sulphur gases in general (and DMS in particular) as sources of NSS—SO₄²⁻ and of CCN in the unpolluted marine troposphere. Because submicrometre NSS—SO₄²⁻ particles are not directly emitted by land or ocean surfaces, they are created in the atmosphere only by chemical reactions. As far as is known, the only significant gaseous sulphur precursors of NSS—SO₄²⁻ of biological origin are DMS from the oceans and H₂S, DMS and perhaps other sulphur species from land biota. These gases are oxidized in air, largely by OH^{15,16,75}, to form SO₄²⁻. At the low NO_x concentrations typical of the unpolluted marine troposphere, the oxidation of DMS by NO₃ can be considered negligible^{17,18}. In addition to SO₂ the oxidation of DMS by OH also produces MSA. At low NO_x

levels, however, the fraction of MSA produced is less than 20%, as suggested both by laboratory experiments (ref. 19 and unpublished manuscript by I. Barnes, V. Bastian and K. H. Becker, 1986) and by the ratio of MSA to NSS—SO₄²⁻ observed on marine aerosols from remote regions^{20,21}. MSA is mostly present as aerosol particles (M.O.A., unpublished data) and due to its high solubility is also likely to be effective as CCN.

The oxidation of DMS to DMSO by the IO radical, which has recently been proposed (unpublished manuscript by I. Barnes *et al.*), is probably not important as a sink for DMS because the concentrations of DMSO in the atmosphere and in marine rain have been found to be very low compared to those of DMS, MSA and NSS—SO₄²⁻ (ref. 22 and M.O.A., unpublished data). The present evidence thus suggests that SO₂ is the major oxidation product of DMS in the unpolluted marine atmosphere. In the presence of cumulus clouds, which are abundant in the marine boundary layer, conversion of SO₂ to NSS—SO₄²⁻ is rapid and significantly exceeds the rate of dry deposition of SO₂ (refs 18, 21).

The rates of DMS oxidation obtained from laboratory experiments, and the corresponding lifetimes, are consistent with observed concentrations, the observed flux out of the atmosphere in rain and the calculated flux into the atmosphere from the ocean surface^{5,17,18}. Thus, the lack of alternative sources of sulphur-bearing gases and the internal consistency of mass-balance calculations support the hypothesis that these biologically derived gases are the probable main sources of NSS—SO₄²⁻ in the remote marine boundary layer. It is therefore reasonable to assume that any change in atmospheric DMS concentration would cause a corresponding change in NSS—SO₄²⁻ concentration and hence in the number of particles which act as CCN. This could be due to changes in the number-concentration of particles, changes of the mass of water-soluble material in existing particles, or both.

Effect on radiative properties of clouds

Clouds of liquid water droplets form only in the presence of CCN. In the unpolluted marine atmosphere, the concentration of particles capable of being CCN varies from about 30 to 200 cm⁻³, depending on aerosol content, supersaturation and meteorological conditions²³. In this same environment, the total number of submicrometre particles is often about 200 cm⁻³, such that a significant fraction of them must be CCN. Bigg²⁴ made observations of CCN at Cape Grim, Tasmania, when the total number-concentration of particles was less than 300 cm⁻³. Results for relative humidities between 100.3 and 101% (values believed to be typical of marine clouds) showed that 40 to 80% of the particles in that marine setting were active CCN. Thus, changes in NSS—SO₄²⁻ in marine air are expected to result in changes in the number-density of cloud droplets, contrary to the widely held belief supported by Fletcher²⁵ that the atmosphere always has an overabundance of particles that could act as CCN. Present-day continental air, by contrast, probably is consistent with Fletcher's generalization.

Koehler²⁶ argued that most CCN are composed of water-soluble materials, such that the marine NSS—SO₄²⁻ should qualify. It is observed that sea-salt particle concentrations at cloud height are typically not more than 1 cm⁻³, so that sea-salt itself cannot be the main CCN^{23,27,28}. Water-soluble materials other than sea-salt and NSS—SO₄²⁻ do exist in marine air but the data are not extensive. Nitrates probably are found on coarser particles, and hence do not contribute substantially to the number population²⁹. Organic compounds exist in the submicrometre particles, but their mass-concentration is probably only a tenth that of NSS—SO₄²⁻ (W. H. Zoller, personal communication). Their number-concentration and sources are unknown.

The idea that natural CCN in marine air consist of sulphates has been believed for decades and that they have a widespread gaseous precursor was suggested by Hobbs²⁷. DMS was iden-

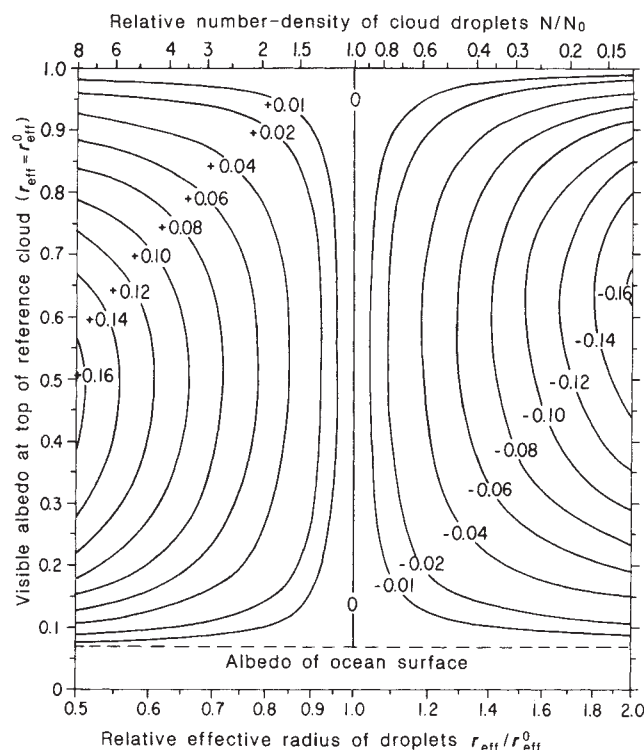


Fig. 1 Change (Δa) of visible albedo ($0.6\text{-}\mu\text{m}$ wavelength) at cloud top caused by changing droplet number-density N while holding vertically integrated liquid water content (liquid water path, LWP) constant. Δa is plotted as a function of the albedo of the reference cloud and the effective radius (r_{eff} , the surface-area-weighted mean radius⁶⁵) of the droplet size distribution relative to that of the reference cloud (r_{eff}^0). The corresponding change in top-of-atmosphere albedo (needed for estimating the effect on Earth radiation budget) can be obtained approximately by multiplying these values by 0.8, as described in footnote of Table 1. The different albedos for the reference cloud (vertical axis) correspond to different values of LWP. When LWP and the shape of the droplet size distribution are held fixed, the number-density of cloud droplets N is related to r_{eff} by $N/N_0 = (r_{\text{eff}}/r_{\text{eff}}^0)^{-3}$, shown on the scale at the top of the figure. These calculations used $r_{\text{eff}}^0 = 8\text{ }\mu\text{m}$, but the figure also is approximately valid for other reference clouds: the plotted values of Δa are in error by less than 20% if albedo < 0.9 and $4 \leq r_{\text{eff}}^0 \leq 500\text{ }\mu\text{m}$ (that is, anywhere in the range of r_{eff} found in real clouds by Hegg⁶⁶). The size distributions used for the calculations are almost monodisperse, broadened just enough to average over the oscillations in the Mie-scattering quantities. However, the calculations are also valid for any realistic size distributions with the same r_{eff} , as Hansen and Travis⁶⁵ showed that the scattering properties of a cloud are controlled essentially by r_{eff} , with very little influence from other moments of size distribution. These calculations assume a direct solar beam at the global average zenith angle $\theta_0 = 60^\circ$, incident on a cloud of spherical droplets of pure water, above an ocean surface. The albedo of an ocean surface under a cloud is essentially independent of wavelength and averages $0.06\text{--}0.08$ (refs 67–69); these calculations assumed 0.07 (dashed horizontal line). The computation of phase function, single-scattering albedo and extinction efficiency for individual cloud droplets used the Mie program of Wiscombe⁷⁰ assuming the refractive index for water is $1.332\text{--}1.09 \times 10^{-8}i$ at $0.6\text{ }\mu\text{m}$ wavelength⁷¹. The computation of radiative transfer in the cloud used the delta-Eddington approximation⁷². This leads to absolute errors in albedo (for water-clouds at visible wavelengths and $\theta_0 = 60^\circ$) of 0.00 to 0.03 depending on cloud optical thickness (Fig. 8 of ref. 73), but the error in albedo differences plotted here is much smaller, generally by a factor of 10.

tified as the likely source of most particles in marine air by Bigg *et al.*³⁰. The size distribution of submicrometre NSS-SO_4^{2-} as deduced from size-resolved samples^{31,32} is about right for activation at supersaturation between 0.1% and 1% which is appropriate for marine clouds and is comparable to the values used by Bigg²⁴. The mass-concentration of NSS-SO_4^{2-} in the remote marine troposphere is about $0.3\text{ }\mu\text{g m}^{-3}$ (refs 31, 33) which, if the number-mean radius is $0.07\text{ }\mu\text{m}$ (which is reasonable⁷⁶) yields a total number-population of about 100 cm^{-3} , in agreement with measured CCN populations. The concentration of NSS-SO_4^{2-} in remote marine rain water is around $2\text{--}5 \times 10^{-6}\text{ M}$ (ref. 34), which agrees with a simple nucleation scavenging calculation with $0.2\text{--}0.5\text{ }\mu\text{g m}^{-3}$ of NSS-SO_4^{2-} aerosol and 1 g m^{-3} liquid water in the cloud³⁵.

Other data also support the idea that NSS-SO_4^{2-} particles are the main contributor to the CCN. Much of the light-scattering aerosol in marine air is volatile at elevated temperatures, evaporating in a fashion identical to sulphates with a range of compositions from H_2SO_4 to $(\text{NH}_4)_2\text{SO}_4$, as deduced from temperature-and-humidity-controlled nephelometry^{76,77}. The droplet-nucleating property of particles in marine air is also heat-labile, with CCN disappearing at $T > 300^\circ\text{C}$ (ref. 36). Finally, the turnover time of CCN from purely physical data in the atmosphere has been deduced to be of the order of one day³⁷, which is the same as results from an estimate for turnover time of NSS-SO_4^{2-} based on mass-concentration, rainfall concentration and rainfall amount, as follows. If we take an NSS-SO_4^{2-} flux, F , of $0.3\text{ g m}^{-2}\text{ yr}^{-1}$ as representative of remote marine sites³⁴, $0.3\text{ }\mu\text{g m}^{-3}$ as the typical concentration, C , of NSS-SO_4^{2-} aerosol, and $3,000\text{ m}$ as its scale height, H , the turnover time, $\tau = HC/F \sim 1$ day.

We now consider the potential effects of NSS-SO_4^{2-} variations on cloud properties. Changing the size distribution or concentration of the CCN causes the size distribution of cloud droplets to change³⁸, which could affect the coalescence and rain production process and possibly the time-averaged cloud cover. However, the effect which is well established (and which we think to be the most significant effect) is that changes in the size distribution of droplets would change the reflectance (albedo) of clouds. This step of the proposed climatic feedback loop is at present more readily quantified than the other steps, so it is presented in some detail. How the average liquid water content of clouds would change is unknown, so we hold it constant in our analysis.

The liquid water content $L(\text{g m}^{-3})$, the number-density of droplets $N(\text{m}^{-3})$, and the droplet radius r (for a monodispersion) are related by

$$L = (4/3)\pi r^3 \rho N$$

where ρ is the density of water in the appropriate units. Various studies have examined the effect of holding one of these three variables constant while the other two change. Paltridge³⁹, Charlack⁴⁰, and Somerville and Remer⁴¹ considered cloud albedo changes due to a climatic warming which might increase the liquid water content of clouds. They held r fixed so that the increase in cloud albedo was due to increase of N . Bohren⁴² showed that the increase of albedo would be somewhat less if N was instead held fixed (that is, assuming that CCN concentration did not change), so that the droplets increased in size rather than number.

For our hypothesis, we must instead consider the effect of holding L constant while increasing N . This leads to a decrease

Table 1 Climatic effect caused by increasing CCN concentration over the ocean

<i>a</i> Global annual average cloud cover (ocean areas only)			<i>b</i> Example: effect on surface climate due to increasing CCN concentration <i>N</i> by 30% while holding liquid water path fixed		
Cloud type*	Ocean area covered (%)	Earth covered by oceanic clouds (%)		For area covered by oceanic stratiform water clouds	Averaged over Earth's surface area
Non-overlapped St/Sc†	25.2	17.6	Imposed change in <i>N</i>	+30%	
Non-overlapped As/Ac‡	10.8	7.5	Change in r_{eff}	−10%	
As/Ac overlapped with St/Sc§	8.8	6.1	Change in 0.5–0.7- μm albedo at TOC #	+0.02	
Nimbostratus, cumulus, cumulonimbus	not applicable (optically thick; high albedo)		Change in 0.5–0.7- μm albedo at TOA**	+0.018	
Cirrus	not applicable (ice)		Change in solar albedo at TOA**	+0.016	+0.005
Total cover of oceanic stratiform water clouds (As/Ac + St/Sc) not overlapped with cumuliform clouds	44.8¶	31.2	Equivalent change in solar constant††		−0.7%
			Change in global-average surface temperature‡‡		−1.3 K

* Only the areas covered by low stratus and stratocumulus (St/Sc) clouds and middle altostratus and altocumulus (As/Ac) are considered, because the change in albedo is smaller for convective clouds and for nimbostratus because they are thicker and may have albedos greater than 0.8.

† The zonal average amount (fractional areal coverage) of St/Sc over the oceans varies from 18% at low latitude to 50% at high latitude⁶¹ and averages 34% for the world ocean, average of all four seasons (ref. 61 and additional data from S.G.W. *et al.*, in preparation). The daytime (sunlit) amount is close to the diurnal average amount. When St/Sc is observed over the ocean, As/Ac is also present about it about 52% of the time⁶². The amount-when-present of oceanic As/Ac is 50% (S.G.W. *et al.*, in preparation), so $50\% \times 52\% = 26\%$ of the St/Sc amount is overlapped by As/Ac. Non-overlapped St/Sc thus covers $34\% \times (1.0 - 0.26) = 25.2\%$ of the ocean area or 17.6% of the Earth's surface.

‡ As/Ac covers 22.4% of the ocean during the daytime (ref. 61 and S.G.W. *et al.*, in preparation) but only 10.8% of the ocean is covered by As/Ac which does not overlap St/Sc, cumulus, or cumulonimbus (using a procedure parallel to that used in the St/Sc analysis above).

§ We assume that this two-layer cloud is still optically thin enough for its albedo to be less than 0.8, that is, in the region of Fig. 1 where Δa is insensitive to *a*.

|| A change in albedo of low or middle water-clouds may be muted at TOA when those clouds are partially hidden by higher ice clouds (cirrus). Here we assume that cirrus is thin enough that the change in planetary albedo due to the water-clouds is the same as if cirrus were absent.

¶ If cumulus is also included, the total areal coverage of clouds whose albedo is sensitive to CCN concentration changes from 44.8% to 56.6%.

From Fig. 1.

** Because of absorption in the 0.6- μm band of ozone above the cloud, the visible-channel albedo at TOA is smaller than at TOC, by a factor of about 0.9 (ref. 47). A further factor of about 0.9 is needed to convert visible-channel TOA albedo to solar TOA albedo^{47,63}, because cloud albedo is lower in the near-infrared than in the visible. The same factors apply to Δa .

†† The global average planetary albedo is now 0.30 (ref. 64), so a change in planetary albedo Δa causes the same change in the amount of solar energy absorbed by the Earth-atmosphere system as would a fractional change in solar constant of $(1.0 - 0.3)\Delta a$.

‡‡ From Table 1 of ref. 49.

in mean radius, which causes an increase in total surface area of droplets in the cloud and thus an increase of cloud albedo. The study of this effect was pioneered by Twomey⁴³. His Figs 12.5 and 12.6 show the albedo, *a*, at cloud top at visible wavelengths as a function of cloud thickness for a reference plane-parallel cloud with uniform droplet radius $r = 8 \mu\text{m}$ ('fairly clean maritime conditions'), as well as the albedo which resulted if *N* was multiplied or divided by 8, causing *r* to decrease or increase respectively by a factor of 2. (Much of Twomey's subsequent work on this topic examined the competing effects on cloud albedo due to absorption of sunlight by dark aerosol particles and the increased number of droplets. Only the latter effect is considered here, because H_2SO_4 and its ammonium salts are transparent in the solar spectrum.) However, his calculations apparently assumed an overhead sun (zenith angle $\theta_0 = 0^\circ$) and a black underlying surface. To study the effects on the global radiation budget we have repeated Twomey's calculations using the global average zenith angle $\theta_0 = 60^\circ$ (which causes the cloud albedo to increase relative to $\theta_0 = 0^\circ$) and assuming an ocean surface as the lower boundary, for many different values of *N*. Figure 1 shows the change in albedo Δa from that of Twomey's reference cloud, caused by variation of N/N_0 in the range 8 to 1/8, where N_0 is the number-density of droplets in the reference cloud. Different thicknesses of the reference cloud are represented on the vertical axis by their albedos. The information in Twomey's Fig. 12.6 is therefore contained in the right and left edges of Fig. 1 here, which also shows the effect of any smaller changes in *N*. Our results agree with those of Twomey where they overlap, showing that the change in albedo due to changing *N* is not sensitive to

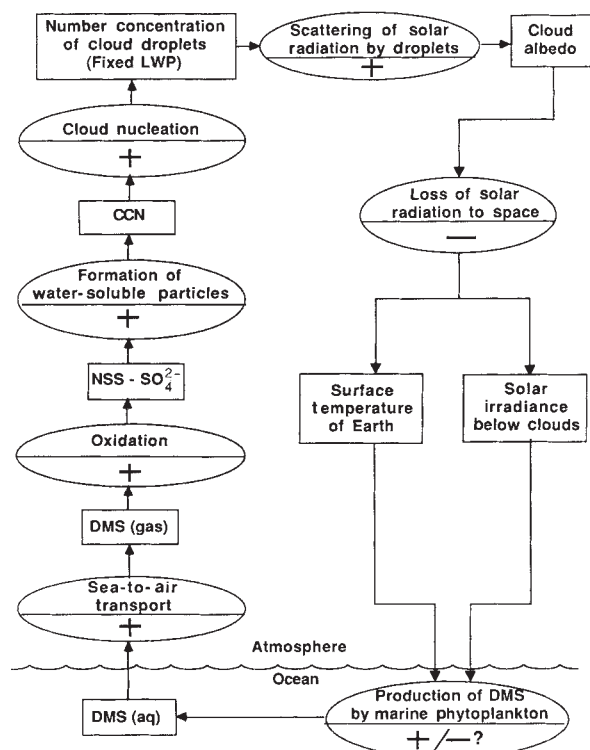
solar zenith angle if the results are expressed as a function of the reference albedo rather than of the cloud thickness. Figure 1 shows that the cloud albedo is most sensitive to *N* at $a = 0.5$ but that Δa is essentially independent of *a* for $0.3 < a < 0.8$, for small changes in N/N_0 . Figure 1 was calculated for effective droplet radius of the reference cloud $r_{\text{eff}}^0 = 8 \mu\text{m}$, but it is approximately valid also for any value of r_{eff}^0 in the range 4–500 μm (see Fig. 1 legend).

Real clouds are not homogeneous, and the albedo of a cloud with a horizontally inhomogeneous distribution of droplets is always less than that of a hypothetical homogeneous cloud^{44,45}, because the albedo-versus-optical-depth function is nonlinear, concave downwards. We therefore compare observed clouds to modelled clouds not by comparing their optical depths but rather by comparing their albedos. (The changes Δa will also be somewhat smaller for an inhomogeneous cloud than for a plane-parallel cloud, but not by much if the albedos of the different cloud elements are all in the range 0.3–0.8, as can be seen from Fig. 1.)

The top-of-atmosphere (TOA) albedo (or 'planetary' albedo) measured by satellites in the 0.5–0.7 μm wavelength channel over marine stratus and stratocumulus (St/Sc) varies in the range 0.25–0.65 (ref. 46). Channel albedo at TOA is usually smaller than channel albedo at top-of-cloud (TOC) by a factor of about 0.9 (ref. 47 and S.G.W. and W. J. Wiscombe, personal communication), because ozone above the cloud absorbs radiation at 0.6 μm , so most marine St/Sc probably have visible TOC albedos, *a*, in the middle region of Fig. 1 where Δa is insensitive to *a* for small relative changes of *N*.

Table 1 illustrates the use of Fig. 1. Stratiform water clouds

Fig. 2 Conceptual diagram of a possible climatic feedback loop. The rectangles are measurable quantities, and the ovals are processes linking the rectangles. The sign (+ or -) in the oval indicates the effect of a positive change of the quantity in the preceding rectangle on that in the succeeding rectangle, following Kellogg's⁷⁴ notation. The most uncertain link in the loop is the effect of cloud albedo on DMS emission; its sign would have to be positive in order to regulate the climate.



that probably have visible TOC albedos in the range 0.3–0.8 cover about 45% of the ocean (Table 1a). As an example (Table 1b), we increase N by 30% over the ocean only (because DMS cannot compete with anthropogenic sulphur over land), which causes r_{eff} to decrease by 10%. From Fig. 1, this causes visible-channel albedo at TOC to increase by 0.02, which causes an increase of 0.016 in planetary albedo averaged over the solar spectrum above these marine clouds, or 0.005 averaged over the entire Earth. (This increase in whole-Earth albedo is smaller than the value 0.006 shown in Fig. 6b of Twomey *et al.*⁴⁸ for this example, because we assume the changes in planetary albedo to apply only to the ocean areas.)

If none of the climatic feedbacks causes cloud albedo to change, the increase in planetary albedo of 0.005 is equivalent to a decrease of the solar constant by 0.7% in a climate model, which causes a decrease of 1.3 K in global mean surface temperature T_s when water-vapour and snow-albedo feedbacks (both positive) are accounted for (Table 1 of ref. 49). This reduction in T_s caused by reducing the effective radius of cloud droplets by 10% everywhere over the world ocean is about one-third as large as the increase in T_s predicted for a doubling of atmospheric CO_2 (ref. 1).

Of course we do not know the relationship between a change in aerosol concentration over the ocean and the resulting change in cloud-droplet concentration N . Theory and experiments (equations 9–1 and 13–41 of ref. 23) indicate that N is proportional to $[\text{CCN}]^p$, with $p \approx 0.8$. Setting $p = 1$ for simplicity, we can show that the cloud-mediated effect on planetary albedo discussed above is much larger than the direct radiative effect of non-nucleated aerosol. Using a mass extinction coefficient of $10 \text{ m}^2 \text{ g}^{-1}$ for $\text{NSS}-\text{SO}_4^{2-}$ (ref. 50), a column-mass of 10^{-3} g m^{-2} , and a ratio of backscattering to total scattering of 0.2, the average total backscattering optical depth, δ_{bsp} , of aerosol particles is 3×10^{-3} , smaller than the Rayleigh backscattering optical depth due to air itself which is ~ 0.1 . This empirically derived quantity is similar to the estimates of Shaw². A 30% increase in the number of particles (over the oceans only) with no change in their mean size would increase δ_{bsp} by 9×10^{-4} , increasing the planetary albedo over dark surfaces by the same amount, but having no effect in areas where clouds are present. The average cloud cover over the oceans is 64% (S.G.W. *et al.*, in preparation)

so the increase in planetary albedo from aerosol alone is $9 \times 10^{-4} \times (1 - 0.64) = 3.2 \times 10^{-4}$ for the ocean or 2.3×10^{-4} for the whole Earth; that is, only about 5% as large as the cloud-mediated increase.

The example we chose for illustration (a 30% change in N) is actually relatively small compared to observed variations. CCN concentrations over the remote ocean can vary with season and time of day, and from one cloud to another, by an order of magnitude or more²³; thus even the extreme left and right sides of Fig. 1 may be applicable for some models of climatic change.

This analysis of the climatic effect of changing the CCN population has assumed that the radiative properties of clouds will change only in the solar spectrum, not in the thermal infrared (beyond $4 \mu\text{m}$ wavelength). Changing the size of droplets will not significantly affect the thermal-infrared emissivity of most water clouds, because they are in effect optically semi-infinite and are nearly black bodies at those wavelengths⁵¹. Cirrus is the only type of cloud which is normally thin enough for its emissivity to be sensitive to optical thickness, but the ice-particle sizes are unlikely to be affected by variations in the concentration of CCN because ice nuclei are normally much rarer and from different origins than CCN.

Global climate and DMS emission

Although we do believe that increased DMS emission should influence both the mass concentration and number population of CCN composed of $\text{NSS}-\text{SO}_4^{2-}$ (and hence cloud albedo), we do not understand quantitatively the relationship of source strength of DMS to CCN number concentration. It seems likely that increased DMS fluxes would increase the CCN population, based for example on the observation in polluted air that gas-to-particle conversion produces new particles⁵².

When looking for feedbacks which link the sea-to-air mass flux F of DMS to global climate (Fig. 2), we can consider the variables which make up the flux equation:

$$F = A \cdot k \cdot \Delta c$$

We can change either the total ocean surface area (A) available for gas exchange, the transfer velocity (k), or the concentration gradient across the air/sea interface (Δc). Because the ocean is

highly oversaturated (by at least three orders of magnitude) relative to the atmosphere, the concentration gradient term is essentially identical to the DMS concentration in surface sea water.

A 10% decrease in ice-free ocean-surface area accompanied the last glacial maximum⁵³. During an ice age, the areas covered by ice on the continents and oceans increase, and there is some exposure of continental shelf now under water. The decrease in ocean area during an ice age could lead to a drop in the global flux of DMS to the atmosphere. Climatic changes would also affect the wind field over the ocean (perhaps only slightly⁵⁴) and thereby the transfer velocity k . The magnitude of this effect on CCN would depend on the relative rates of loss of DMS to the atmosphere and in the water column.

Empirically, we find that the largest flux of DMS comes from the tropical and equatorial oceans⁵. This suggests that the most important climatic role of DMS is to contribute to elevated cloud albedo over the warmest ocean regions, and thus to reduce the input of heat into the low-latitude oceans. A cooling of the oceans or a reduction in area of the tropical seas could thus lead to a smaller DMS flux, providing a stabilizing negative feedback.

On the other hand, the sea-to-air flux of DMS and consequently the albedo of marine clouds could significantly change as a result of ecological changes which would favour phytoplankton species with large DMS output rates over those with low output rates, or vice versa. At this time we have only examined a few species for their DMS emission rates. Although we know that there are large interspecific differences, we cannot yet relate them to phytoplankton taxonomy. In particular, we are not yet able to use the data on phytoplankton speciation during the geological past, which has been obtained by the CLIMAP program⁵³, to predict the DMS flux during periods of glaciation. As a result, we are left with an incomplete story: we are convinced that the emission of DMS into the marine atmosphere plays a crucial climatic role, but we cannot yet define precisely the processes which regulate the rate of DMS emission.

Geophysiology and homeostasis

Gaia theory³ suggests that in order to maintain thermostasis on Earth, CO₂ is continuously and increasingly pumped from the atmosphere. There is a constant input from tectonic processes, and the long-term sink is the burial of carbonate rock in the sediments. The sink for CO₂ is almost wholly biologically determined; without life, CO₂ would rise to an abundance of well over 1% by volume. Lovelock and Whitfield⁵⁵ observed that if climate regulation does take place by pumping of CO₂ then the mechanism is now close to the limit of its capacity to operate. Atmospheric CO₂ has been reduced from about 30% of the atmosphere at the start of life to the present 300 p.p.m.v., a factor of 1,000. It was suggested that the decrease of CO₂ through its declining greenhouse effect had compensated for the monotonic increase of solar luminosity and so the climate had remained constant and suited for life. It cannot be much more reduced without seriously impairing the growth of mainstream plants whose limit is near 150 p.p.m.v.; not much less than the minimum, 180 p.p.m.v., of the last glaciation⁷⁸.

DMS emission, through its effect on the planetary albedo, shares with CO₂-pumping a cooling tendency. So if the DMS production increases with temperature and/or solar irradiance, the sign of its climatic effect would be in the right direction to offset what seems, for the biota, to be an excessive solar flux.

How could the local activity of species living in the ocean evolve to serve in the altruism of planetary regulation? The biogeochemical cycles of carbon and of sulphur are intimately linked and appear to be connected with the regulation of redox potential in both oxic and anoxic ecosystems^{56,57}. The first indication of a geophysiological role for sulphur came from the observations of Challenger⁵⁸ that marine algae emitted DMS. He found that some species of shoreline algae of the order

Polysiphonia contained as much as 15% of their dry weight as the sulphur betaine, DMSP. Until recently it has been difficult to envisage any geophysiological link between DMS production in the oceans or on the shoreline and the need of land-based ecosystems for sulphur. Why should an algal community of the ocean make the extravagant altruistic gesture of producing DMS for the benefit of, among other things, elephants and giraffes? One possible answer is that the biosynthesis of DMS began as a local activity that grew to become an unconscious benefit for the system.

With algae it seems likely that the local problem that led to the biosynthesis of DMSP was salt stress. As a result of tidal movement algae are left exposed and drying on the beach at least twice daily. Neutral solutes, such as glycerol or dimethyl sulphoxide are well known to protect against the adverse effects of freezing or drying. The mechanism of the protective effect is simply the solvent effect for salt of these involatile compounds⁵⁹. Among solutes able to protect cells against desiccation are the betaines. These compounds although ionic are charge-neutral; the negative and positive moieties are on the same molecule. DMSP, (CH₃)₂S⁺·CH₂CH₂COO⁻ (a thionium betaine), is widely distributed among marine algae and serves to protect them against drying or increased salinity. Thus the requirement of an osmoregulatory substance by intertidal organisms may have been the original reason for DMSP synthesis.

We begin to see a possible geophysiological link between the local self interest of salt-stress prevention and the global sulphur cycle. The accidental by-product of DMSP production is its decomposition product DMS. This compound or its aerosol oxidation products will move inland from the shore and deposit sulphur over the land surface downwind of the ocean. The land tends to be depleted of sulphur and the supply of this nutrient element from the ocean would increase productivity and the rate of weathering and so lead to a return flow of nutrients to the ocean ecosystems. What seems a naive altruism is in fact an unconscious self-interest. Sulphur from DMS can travel farther than the sea-salt aerosol because several steps are involved in the conversion of gaseous DMS to aerosol sulphate; also the resulting aerosol particles are smaller and so have much longer lifetimes.

A large proportion of the current biosynthesis of DMSP is in the open oceans distant from the land surfaces. Is this DMSP also made for the relief of salt stress, or is it a redundant mechanism kept in action because of glacial epochs when the sea or part of it was saltier? Interglacials have occupied only one tenth of the time during the current series of glaciations; this may be too short a period for the devolution of DMSP biosynthesis. Alternatively, it may be that production of DMSP in the open ocean has a different geophysiological basis from that in the continental shelf regions and one that is unconnected with salinity as such.

We have seen how the local self-interest of shoreline algae could lead to the mutual sharing of sulphur and nutrients with land-based organisms. The evolution of a link between ocean climate and DMS production could have happened in a similar way. Ocean organisms are often deficient in nitrogen and to some extent vulnerable to solar ultra-violet. Cloud formation with rainfall would return nitrogen to the ocean and also serve as a sunshade. If either or both of these effects were significant for the health of phytoplankton then species that emitted DMS might be at an advantage.

Is the sulphur cycle also involved in global climate control? Evapotranspiration is known to modify the climate of forests in the humid tropics. The additional cloud cover that comes from the vast water vapour flux of the trees increases the planetary albedo and further cools the surface. The emission of DMS from the oceans seems to act similarly. Through its aerosol oxidation products it alters the properties of clouds, increasing the albedo of the oceanic regions and hence of the greater part of the planet. The link between the biota and climate in both

of these processes of cloud formation could be a mechanism for climate control, the clouds serving as do white daisies in the 'Daisyworld' model of Gaian climate regulation⁶⁰.

Lastly, DMS is the principal component of the present biogeochemical flux of sulphur to the atmosphere. But it is not the only one; some sulphur is emitted as H₂S, COS and CS₂. COS, both from direct emission and as an oxidation product of CS₂, is stable in the troposphere long enough to be an important source of sulphur to the stratosphere. COS in the stratosphere would be oxidized and produce a sulphate aerosol there. Such stratospheric aerosols scatter sunlight back to space and lead to a cooler climate. The biological variation of COS output is therefore another possible geophysiological means of regulating climate.

Future research needs

We propose that sulphate aerosols derived from the sulphur gases produced by the marine biota are important determinants of cloud albedo and, as a consequence, the climate. It also seems likely that the rate of DMS emission from the oceans is affected by the climate, thus closing a feedback loop.

There are significant gaps in our knowledge of this proposed feedback system. Most importantly, we need to understand the climatic factors affecting DMS emission. Because some species produce much more DMS than others, we must include the necessary understanding of controls on phytoplankton species

abundance. We also need to understand the relationship between DMS concentration in the air and the CCN population, through the intervening aerosol physical processes. Knowing how the area of cloud cover is influenced by CCN is also important.

Nonetheless, the role of the CCN population in controlling albedo, the production of CCN in marine air by the oxidation of DMS from the biota and the sensitivity of the Earth's temperature to the CCN population seem to be established. Although we do not understand the details of the climatic feedback, it seems that CCN from biogenic DMS currently act to cool the Earth. It is possible that the Earth's climate has been mediated in the past (for instance, that this feedback has helped to counteract the increasing luminosity of the Sun and/or that it has already counteracted the influence of the recent increase in CO₂ and other 'greenhouse' gases). However, the data required to demonstrate the latter effect have not been and are not now being acquired.

The portion of this work done in the USA was supported in part by NSF grants ATM-82-15337, ATM-83-18028, OCE-83-15733 and ATM-84-07137. The computations were done at the National Center for Atmospheric Research. We thank Marcia Baker, Keith Bigg, Robert Chatfield, Ian Galbally, Dean Hegg, Ann Henderson-Sellers, Kendal McGuffie, Henning Rodhe and Stanley Thompson for commenting on an early draft, and Antony Clarke for discussions.

- Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109-115 (1986).
- Shaw, G. *Climatic Change* **5**, 297-303 (1983).
- Lovelock, J. E. *Bull. Am. met. Soc.* **67**, 392-397 (1986).
- Andreae, M. O. in *The Biogeochemical Cycling of Sulfur and Nitrogen in the Remote Atmosphere* (eds Galloway, J. N., Charlson, R. J., Andreae, M. O. & Rodhe, H.) 5-25 (Reidel, Dordrecht, 1985).
- Andreae, M. O. in *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 331-362 (Reidel, Dordrecht, 1986).
- Cullis, C. F. & Hirschler, M. M. *Atmos. Environ.* **14**, 1263-1278 (1980).
- Andreae, M. O. in *Microbial Mats: Stromatolites* (eds Cohen, Y., Castenholz, R. W. & Halverson, H. O.) 455-466 (Liss, New York, 1984).
- Andreae, M. O. & Andreae, T. W. *J. geophys. Res.* (submitted).
- Vairavamurthy, A., Andreae, M. O. & Iverson, R. L. *Limnol. Oceanogr.* **30**, 59-70 (1985).
- Turner, S. M. & Liss, P. S. *J. Atmos. Chem.* **2**, 223-232 (1985).
- Bates, T. S., Cline, J. D., Gammon, R. H. & Kelly-Hansen, S. R. *J. geophys. Res.* (in the press).
- Andreae, M. O. *Limnol. Oceanogr.* **30**, 1208-1218 (1985).
- Brimblecombe, P. & Shooter, D. *Mar. Chem.* **19**, 343-353 (1986).
- Andreae, M. O., Barnard, W. R. & Ammons, J. M. *Ecol. Bull. (Stockholm)* **35**, 167-177 (1983).
- Niki, H., Maker, P. D., Savage, C. M. & Breitenbach, L. P. *Int. J. chem. Kinet.* **15**, 647-654 (1983).
- Yin, F., Grosjean, D. & Seinfeld, J. H. *J. geophys. Res.* (in the press).
- Ferek, R. J., Chatfield, R. B. & Andreae, M. O. *Nature* **320**, 514-516 (1986).
- Andreae, M. O. *et al. J. geophys. Res.* **90**, 12891-12900 (1985).
- Tyndall, G. S., Burrows, J. P., Schneider, W. & Moortgat, G. K. *Chem. Phys. Lett.* **130**, 463-466 (1986).
- Saltzman, E. S., Savole, D. L., Prospero, J. M. & Zika, R. G. *J. Atmos. Chem.* **4**, 227-240 (1986).
- Andreae, M. O., Berresheim, H., Andreae, T. W., Kritz, M. A., Bates, T. S. & Merrill, J. T. *J. Atmos. Chem.* (in the press).
- Harvey, G. R. & Lang, R. F. *Geophys. Res. Lett.* **13**, 49-51 (1986).
- Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* (Reidel, Dordrecht, 1978).
- Bigg, E. K. *Atmos. Res.* **20**, 82-86 (1986).
- Fletcher, N. H. *The Physics of Rainclouds* 98 (Cambridge University Press, 1962).
- Koehler, H. *Trans. Faraday Soc.* **32**, 1152 (1936).
- Hobbs, P. V. *Q. Jl R. met. Soc.* **97**, 263-271 (1971).
- Radke, L. F. *thesis*, Univ. Washington, Seattle (1968).
- Charlson, R. J., Chameides, W. & Kley, D. in *The Biogeochemical Cycling of Sulfur and Nitrogen in the Remote Atmosphere* (eds Galloway, J. N., Charlson, R. J., Andreae, M. O. & Rodhe, H.) 67-80 (Reidel, Dordrecht, 1985).
- Bigg, E. K., Gras, J. L. & Evans, C. J. *J. Atmos. Chem.* **1**, 203-214 (1984).
- Andreae, M. O. *J. geophys. Res.* **87**, 8875-8885 (1982).
- Savoie, D. L. & Prospero, J. M. *Geophys. Res. Lett.* **9**, 1207-1210 (1982).
- Prospero, J. M., Savoie, D. L., Nees, R. T., Duce, R. A. & Merrill, J. J. *J. geophys. Res.* **90**, 10586-10596 (1985).
- Galloway, J. N. in *The Biogeochemical Cycling of Sulfur and Nitrogen in the Remote Atmosphere* (eds Galloway, J. N., Charlson, R. J., Andreae, M. O. & Rodhe, H.) 143-175 (Reidel, Dordrecht, 1985).
- Charlson, R. J. & Rodhe, H. *Nature* **295**, 683-685 (1982).
- Twomey, S. A. *J. Atmos. Sci.* **28**, 377 (1971).
- Wallace, J. M. & Hobbs, P. V. *Atmospheric Science, an Introductory Survey* 212 (Academic, New York, 1977).
- Hobbs, P. V., Harrison, H. & Robinson, E. *Science* **183**, 909-915 (1974).
- Paltridge, G. W. *Q. Jl R. met. Soc.* **106**, 895-899 (1980).
- Charlock, T. P. *Tellus* **34**, 245-254 (1982).
- Somerville, R. C. J. & Remer, L. A. *J. geophys. Res.* **89**, 9668-9672 (1984).
- Bohren, C. F. *J. geophys. Res.* **90**, 5867 (1985).
- Twomey, S. *Atmospheric Aerosols* (Elsevier, Amsterdam, 1977).
- Welch, R. M. in *Fifth Conference on Atmospheric Radiation* 505-507 (American Meteorological Society, Boston, 1983).
- Gabriel, P., Lovejoy, S., Austin, G. & Schertzer, D. in *Sixth Conference on Atmospheric Radiation* 230-235 (American Meteorological Society, Boston, 1986).
- Coakley, J. A. Jr & Davies, R. J. *J. Atmos. Sci.* **43**, 1025-1035 (1986).
- Wiscombe, W. J., Welch, R. M. & Hall, W. D. *J. Atmos. Sci.* **41**, 1336-1355 (1984).
- Twomey, S. A., Piepgrass, M. & Wolfe, T. L. *Tellus* **36B**, 356-366 (1984).
- Wetherald, R. T. & Manabe, S. *J. Atmos. Sci.* **32**, 2044-2059 (1975).
- Waggoner, A. P. *et al. Nature* **261**, 120-122 (1976).
- Paltridge, G. W. & Platt, C. M. R. *Radiative Processes in Meteorology and Climatology* 191 (Elsevier, Amsterdam, 1976).
- McMurry, P. H. & Wilson, J. C. *J. geophys. Res.* **88**, 5101-5108 (1983).
- McIntyre, A. *et al. Science* **191**, 1131-1144 (1976).
- Bryan, K. & Spelman, M. J. *J. geophys. Res.* **90**, 11679-11688 (1985).
- Lovelock, J. E. & Whitfield, M. *Nature* **296**, 561-563 (1982).
- Garrels, R. M. & Lerman, A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4652-4656 (1981).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton University Press, 1984).
- Challenger, F. *Adv. Enzymol.* **12**, 429-491 (1951).
- Lovelock, J. E. *Biophys. Acta* **11**, 28-34 (1953).
- Watson, A. J. & Lovelock, J. E. *Tellus* **35B**, 284-289 (1983).
- Warren, S. G., Hahn, C. & London, J. in *Fifth Conference on Atmospheric Radiation* 313-314 (American Meteorological Society, Boston, 1983).
- Warren, S. G., Hahn, C. & London, J. *J. Clim. appl. Meteor.* **24**, 658-667 (1985).
- Shine, K. P., Henderson-Sellers, A. & Slingo, A. *Q. Jl R. met. Soc.* **110**, 1170-1179 (1984).
- Stephens, G. L., Campbell, G. G. & Vonder-Haar, T. H. *J. geophys. Res.* **86**, 9739-9760 (1981).
- Hansen, J. E. & Travis, L. D. *Space Sci. Rev.* **16**, 527-610 (1974).
- Hegg, D. A. *J. Atmos. Sci.* **43**, 399-400 (1986).
- Briegleb, B. & Ramanathan, V. *J. appl. Met.* **21**, 1160-1171 (1982).
- Kondratyev, K. *Radiation in the Atmosphere* (Academic, New York, 1969).
- Payne, R. E. *J. Atmos. Sci.* **29**, 959-970 (1972).
- Wiscombe, W. J. *Appl. Opt.* **19**, 1505-1509 (1980).
- Hale, G. M. & Querry, M. R. *Appl. Opt.* **12**, 555-563 (1973).
- Joseph, J. H., Wiscombe, W. J. & Weinman, J. A. *J. Atmos. Sci.* **33**, 2452-2459 (1976).
- King, M. D. & Harshvardhan, J. *Atmos. Sci.* **43**, 784-801 (1986).
- Kellogg, W. W. in *Climate of the Arctic* (eds Weller, G. & Bowling, S. A.) 111-116 (Geophysical Institute, Fairbanks, Alaska, 1975).
- Graedel, T. E. *Geophys. Res. Lett.* **6**, 329-331 (1979).
- Clarke, A. D., Ahlquist, N. C. & Covert, D. S. *J. geophys. Res.* **92** (in the press).
- Covert, D. S. *J. geophys. Res.* (submitted).
- Nefel, A., *et al. Nature* **295**, 220-223 (1982).

Genome organization and transactivation of the human immunodeficiency virus type 2

Mireille Guyader, Michael Emerman, Pierre Sonigo, François Clavel*, Luc Montagnier & Marc Alizon†

Unité d'Oncologie Virale (CNRS UA 1157), and Unité de Recombinaison et Expression Génétique (INSERM U163, CNRS UA271), Laboratoire de Biologie Moléculaire et Immunologie des Rétrovirus, Institut Pasteur, 25 rue du Dr Roux 75724 Paris Cedex 15, France

Analysis of the nucleotide sequence of the human retrovirus associated with AIDS in West Africa, HIV-2, shows that it is evolutionarily distant from the previously characterized HIV-1. We suggest that these viruses existed long before the current AIDS epidemics. Their biological properties are conserved in spite of limited sequence homology; this may help the determination of the structure-function relationships of the different viral elements.

THE acquired immune deficiency syndrome (AIDS) has now spread worldwide and appears to be an acute public health problem in Africa in particular¹⁻⁵. A retrovirus designated human immunodeficiency virus (HIV), but previously known as LAV, HTLV-III or ARV, was shown to cause AIDS in the different areas afflicted by the epidemics⁶⁻⁸. Indeed, isolates from North America, Western Europe and Central Africa have the same biological properties, and antigenically cross-reactive proteins with the same relative molecular mass⁹⁻¹¹. Only studies at the molecular level have revealed some differences in the nucleotide sequence of North-American and African isolates^{12,13}. This sequence variation is also present, though to a lesser extent, among different isolates from the USA¹⁴⁻¹⁸.

The western part of Africa seemed relatively spared by AIDS³. Recently, however, several typical cases were found in a survey of patients from Guinea Bissau and other countries of West Africa¹⁹⁻²¹. Unexpectedly, most of these patients did not have detectable titres of antibodies against HIV. But they were found to be infected by a retrovirus related to HIV by its ultrastructural and biological properties, such as cytopathogenicity and tropism for cells carrying the CD4(T4) antigen¹⁹. Antibodies raised against HIV could immunoprecipitate the *gag* and *pol* products of these isolates, which have molecular masses that are similar but not identical to these antigens of HIV; in contrast, the *env* products could not be immunoprecipitated, whereas previous HIV isolates showed wide cross-antigenicity of the envelope glycoprotein. Furthermore, the genome of this new retrovirus cross-hybridized only poorly in very low stringency conditions with HIV DNA probes^{19,22}. We have therefore designated this West African AIDS virus as HIV type 2 (HIV-1 referring to the AIDS retrovirus previously identified in Central Africa, North America and Europe). More than 20 isolates have so far been made from patients with AIDS and related conditions, mainly originating from west Africa^{20,21}, but also in some Europeans (L.M., unpublished), and epidemiological studies in progress indicate a seroprevalence of 1-2% in some populations of West Africa (F. Brun-Vézinet, personal communication).

HIV-2 appears to be closely related to the simian immunodeficiency viruses (SIV) a group of cytopathic retroviruses whose prototype, STLV-3_{mac}, was identified in captive rhesus monkeys (*Macaca mulatta*) with an AIDS-like disease²³, and was later found to infect other primate species, either wild or in captivity²⁴⁻²⁶. Genetic comparisons of SIV, HIV-1 and HIV-2 may help to elucidate the phylogeny of these viruses and the origins of the recent AIDS epidemics. As these retroviruses share most of their biological properties, the identification of conserved

sequences is important to localize the functional domains of the viral proteins and regulating elements, and design new diagnostic and therapeutic tools. We present here the complete nucleotide sequence of HIV-2, the comparison of its proteins with those of HIV-1, and preliminary studies on the regulation of HIV-2 expression.

Nucleotide sequence and LTR analysis

The sequence presented in Fig. 2 is derived from two λ clones corresponding to integrated proviral DNA from the ROD isolate of HIV-2 (ref. 22), obtained in 1985 from an AIDS patient from Cape Verde Islands (offshore Senegal, refs 19, 20). The genome of HIV-2 is 9,671 nucleotides long (in its RNA form), whereas HIV-1 isolates are about 9,200 nucleotides long. This difference is partly explained by the respective sizes of the long terminal repeats (LTRs, see below).

The genetic organization of HIV-2 (shown in Fig. 1) is analogous to that of HIV-1, that is:

5'LTR-*gag-pol*-central region-*env*-orf F-3'LTR.

The 'central region', also identified in the ovine lentivirus visna²⁷, contains five major open reading frames (ORFs), four being clearly related to the ORFs of HIV-1 that encode the Q (or *sor*), R, *tat* and *art* (or *trs*) genes of HIV-1 (refs 15-18, 27-31). The fifth, which we designate ORF X, has no obvious counterpart in HIV-1. Alignments of the nucleotide sequences of HIV-1 and 2 show their distant homology (from ~60% for the more conserved *gag* and *pol* genes, to 30-40% for the other viral genes and LTRs). To allow these alignments to be made many insertions and deletions must be introduced into the sequences. We do not find that these insertions are the small duplications that would be characteristic of the recent divergence of retroviral sequences, as was noted among isolates of HIV-1 (ref. 12).

The limits of the LTRs and of their internal U3, R and U5 elements, determined by sequence analysis and some complementary experiments, are shown in Fig. 2. Classically bounding the retroviral LTRs are short inverted repeats (5' CTG-CAG 3') located after a polypurine tract for the 3'LTR, and before a sequence complementary to the 3' end of a transfer RNA that is used as primer by the reverse transcriptase (here, as in HIV-1 and visna virus, a lysine tRNA, refs 15, 27) for the 5' LTR. The R-U5 junction, corresponding to the 3' end of the polyadenylated viral RNA, was previously localized by sequencing oligo(dT)-primed complementary DNA (cDNA) derived from the HIV-2_{ROD} genome²². The length of U5 + R, and hence the position of the U3-R junction corresponding to the 5' cap site of the viral RNA were deduced from the size of a HIV-2 cDNA synthesized using the endogenous reverse transcriptase activity and the endogenous tRNA^{lys} primer (see Fig. 3). This

* Present address: Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA

† To whom correspondence should be addressed.

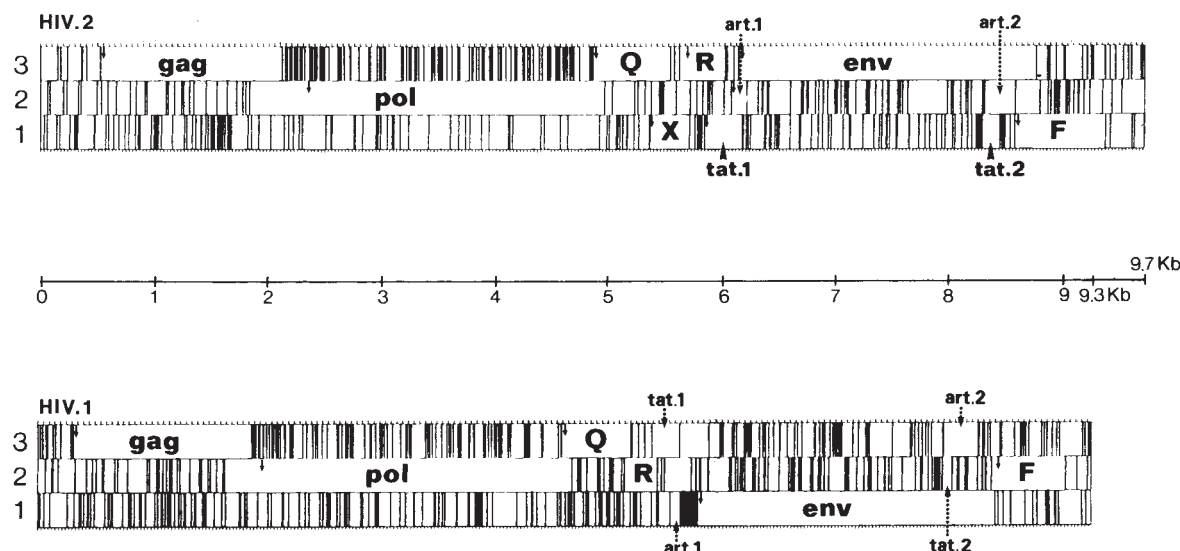


Fig. 1 Organization of the HIV-2 and HIV-1 genome (BRU isolate, ref. 15). Vertical bars represent the stop codons in the 3 reading frames. Arrows indicate the initiator AUG codons in viral genes or potential genes. *Tat* 1 and 2, *art* 1 and 2 are the open reading frames containing the coding exons of the *tat* and *art* genes.

'strong-stop cDNA' is 302 $\pm$ 1 nucleotides long (181 nucleotides in HIV-1, ref. 15). Thus, the U5 element is 125 bp long, U3 is 556 bp and R 173 bp (respectively 82, 456 and 97 bp in HIV-1). All the elements of the HIV-2 LTRs are larger than in HIV-1, and alignment by computer programs shows large insertions and very distant overall homology for the aligned regions²². However, the three Sp1 binding sites identified in HIV-1 (ref. 32), are also present in HIV-2 from nucleotide 9,419 to 9,448 with 17 out of 29 nucleotides homologous to this region of HIV-1. The core enhancers identified in HIV-1 (ref. 33) are present in HIV-2 from nucleotide 9,389 to 9,416: the first is 50% homologous and the second 100% homologous to that in HIV-1 (Fig. 2).

The analysis of the virus-specific poly(A)⁺ RNA (not shown) from a cell line infected with and continuously producing HIV-2 revealed a pattern of transcription reminiscent of that observed in HIV-1-infected cells: RNA of over 9 kilobases (kb), corresponding to a full-length transcript, and three types of spliced messenger RNA of 5, 4.5 and 2 kb, also observed in HIV-1 (refs 18, 34).

The *gag* and *pol* proteins and HIV phylogeny

The *gag* precursor of HIV-2 has a calculated relative molecular mass of 57,100 (M_r 57.1K), consistent with the p55 antigen²⁰ seen by immunoprecipitation with patient sera, and is probably processed, by analogy with HIV-1, into the proteins designated p16, p26 and p12 (refs 19, 20). By analogy with the p18^{gag} of HIV-1, p16 would be at the amino terminus of *gag* and precede p26, whose amino terminus has been sequenced (H. Marquardt, personal communication) and starts with the proline residue at position 951. The carboxy-terminal part of the *gag* precursor encodes a p12 that contains the cysteine-rich consensus of the retroviral nucleic-acid-binding proteins also found twice in the p13^{gag} of HIV-1 (ref. 15). The HIV-2 *pol* ORF could encode the p64 and p36 antigens of HIV-2 (ref. 20) which by analogy correspond to the p68 and p34 (reverse transcriptase and endonuclease, respectively³⁵) of HIV-1.

The *gag* and *pol* proteins of HIV-1 and 2 were expected to share large conserved domains, as these HIV-2 proteins can be precipitated by antibodies in sera from patients infected with HIV-1. However, we found that only 58% and 59.4% of the amino acids of *gag* and *pol* respectively are identical to the

corresponding HIV-1 products (Table 1a), whereas the more distant isolates of HIV-1 (Zairian and US) show 90 to 95% amino-acid identity in these proteins (Table 1b and ref. 12). Several insertions and deletions have to be introduced in the alignments (data not shown), whereas they are rare in the comparisons of *gag* and *pol* genes between HIV-1 isolates. The *gag* and *pol* proteins of HIV-2 are no closer to those of the Zairian isolates than to the prototype HIV-1 (BRU isolate) isolated in 1983 from a French patient⁶ probably infected in the USA. Overall, the difference in *gag* and *pol* between HIV-1 and HIV-2 is of the same order as that observed among the group of the human T-cell leukaemia viruses (HTLV-I and II) and bovine leukaemia virus (BLV). However, this latter group displays a higher conservation in the envelope, 70% amino-acid identity between HTLV-I and HTLV-II, versus about 42% between HIV-1 and HIV-2 (see below). Alignments of different retroviral *pol* proteins (Table 1b) confirm that the HIVs form a subgroup that is more related to the lentiviruses visna and equine infectious anaemia virus (EIAV) than to any other human or animal retrovirus.

Homologous domains in *env*

The envelope glycoproteins of retroviruses are translated from a subgenomic viral mRNA (here probably the transcript of 4.5 kb). Addition of sugar residues (*N*-linked glycosylation) gives rise to a high- M_r precursor which is processed by proteolytic cleavage. The length of the leader sequence of the HIV-2 glycoprotein cannot be precisely determined by alignment with that of HIV-1 (experimentally found to be 32 amino acids long³⁶) because of a lack of sequence homology (Fig. 4). But the amino terminus of *env* contains a relatively hydrophobic stretch in the calculated hydropathy plot (not shown) that is probably the signal peptide. The potential cleavage site between the external envelope glycoprotein (120K) and the transmembrane protein (previously thought to be the 36K antigen¹⁹, and now putatively identified as a 40K antigen²⁰) is found at amino acid 505 (Fig. 4) immediately after the Lys-Glu-Lys-Arg sequence. This cleavage site aligns partly to one (Lys-Ala-Lys-Arg) of the two potential cleavage sites found in HIV-1 (the other being located after the Arg-Glu-Lys-Arg stretch). The calculated M_r of the extracellular glycoprotein (EGP) and of the transmembrane protein (TMP) of HIV-2 would be 57K and 41.7K respectively; the discrepancy

Fig. 2 Complete nucleotide sequence of the HIV-2_{ROD} provirus genome. The sequence of the 9,671 nucleotide coding strand from the cap site to the polyadenylation site of the RNA is shown together with the deduced amino-acid sequence of the viral proteins. The 3-bp inverted repeats bounding the LTR are underlined with arrows. The primer binding site (PBS) complementary to the 3' end of tRNA^{lys}, and polypurine tract (PPT) are underlined. The repeat of the PPT in the *pol* gene is also underlined. The limits of the U3, R and U5 elements of the LTR are indicated by arrows. The promoter (TATAAA), the three potential Sp1 factor binding sites (indicated by dotted lines) and the two core enhancer sequences (E, indicated by dashed lines) of U3 are shown; in R the polyadenylation signal AATAAA is underlined. The viral genes and potential genes are translated from the first AUG of the open reading frame (ORF), except *pol* which is translated from the beginning of the ORF. The end of translation is indicated by a filled circle. The amino terminus of the major core protein, the p26^{gag} was determined by protein sequencing. SD at position 6,140 and SA at position 8,307 indicate the probable splice donor and acceptor sites of the intron separating the two coding exons of the *tat* and *art* genes.

Methods. The sequence was determined by the M13 shotgun cloning and dideoxynucleotide chain terminator method^{57,58}, as described²⁷ starting from the inserts of the λ phages ROD 27 and ROD 35 containing integrated proviral DNA²². ROD 27 corresponds to the 5' part of the genome, to the *EcoRI* site at position 2,658, whereas ROD 35 corresponds to the part of the genome 3' to this site. An in-frame TAG stop codon was found in ROD 35 *env* gene. The sequencing of the corresponding region in an oligo(dT)-primed HIV-2 cDNA (clone E2, described in ref. 22) revealed that it was due to a C-to-T mutation at position 8,304.

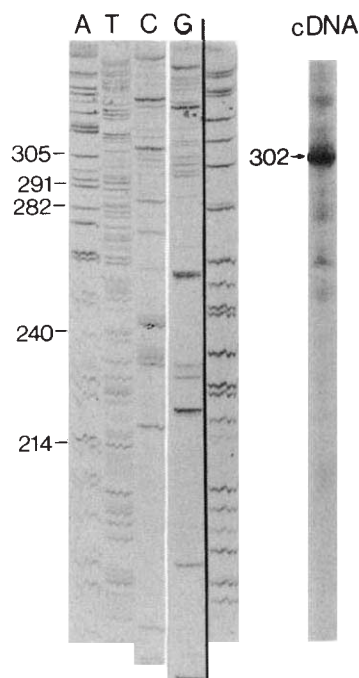


Fig. 3 HIV-2 strong-stop cDNA corresponding to the length of the R+U5 elements of the HIV-2 LTR. The methods were previously described¹⁵. Briefly, virions were purified by ultracentrifugation and an endogenous cDNA reaction performed with radiolabelled nucleotides after mild disruption of viral envelope with Triton X-100. The tRNA^{lys} primer, complementary to the PBS site flanking the U5 element at the 5' end of the genome, was then degraded by alkaline hydrolysis and the cDNA run on a denaturing 6% acrylamide-urea gel together with a sequence reaction for accurate estimation of the size of the products.

with the apparent M_r of the EGP is explained by glycosylation (30 sites in HIV-2, about half of which are conserved with respect to HIV-1).

Figure 4 shows an alignment of the envelopes of the two HIVs. The proteins are overall very distantly related (41.7% identity in the entire envelope, 39.4% in the EGP, 44.8% in the TMP) compared to divergent isolates of HIV-1 (about 75–80% identity in the whole envelope, ref. 12). Many large insertions have to be introduced, particularly in alignment of the EGPs where only short, widely separated domains are conserved between HIV-1 and 2. These domains are clustered into the conserved regions of the EGP of HIV-1 (identified by comparison of different isolates^{12–14}), and generally coincide with cysteine residues. Among the HIV-1 isolates, all the cysteine residues could be aligned in spite of the generally large genetic variation, especially in gp110. Almost all (22/23) of the cysteine residues of HIV-1 can also be aligned with HIV-2, but the latter contains seven additional cysteine residues, often in the regions representing insertions relative to HIV-1. Thus, the folding of the HIV-2 EGP could be different from that of HIV-1, and some regions, therefore, might be exposed in a different manner.

Other viral proteins

The HIV-1 genome contains several other genes encoding proteins of small M_r (10 to 27K), two of which (*tat* and *art/trs*) have an identified function: the positive regulation of viral expression^{30–33}. No role has yet been identified for the p23 encoded by ORF Q (or *sor*)^{37,38}, nor for the p27 encoded by ORF F (or 3' ORF)³⁹. We also observed in the region between the *pol* and *env* genes of HIV-1 (central region) another potential gene, which we designated R (ref. 12). All these elements are found in HIV-2, but the corresponding proteins are only distantly homologous (see Table 1a). In the F protein, most of the difference between HIV-1 and 2 is due to a large insertion in

the amino terminus of HIV-2. The second half of the protein, encoded by the U3 element of the LTR, shows better conservation (data not shown).

Based upon sequence homologies with HIV-1, the *tat* and *art* genes of HIV-2 are probably organized as split genes transcribed into ~2 kb mRNA made of three exons^{18,28–31}: the 5' leader, a first coding exon located in the central region and probably ending at a possible splice donor found at position 6,140 (CAAGT, Fig. 2), and a last exon probably starting at the splice acceptor at position 8,307 in HIV-2 (CAGATC). The *tat* protein of HIV-2 would be longer than that of HIV-1 (130 versus 86 amino acids), having two large insertions in the amino terminus and in the second coding exon (Fig. 4). The main domain of homology of the *tat* proteins corresponds to a region very rich in cysteine residues whose structure is reminiscent of that of the 'cysteine fingers' of some transcription-regulating elements that interact with nucleic acids, such as the TFIIIA factor⁴⁰. This region is followed by an Arg-Lys-rich stretch that could also interact with DNA or RNA. No significant homology is seen in the second coding exon, which has been shown to be dispensable to the function of the protein^{28,29}. The *art*-encoded protein is shorter in HIV-2 than it is in HIV-1 (100 versus 116 amino acids), and most of its length is encoded by the last exon. The most homologous part is located in a stretch of basic residues that may be able to interact with nucleic acids.

Cross-transactivation of HIV-1 and HIV-2

The trans-activator gene (*tat*) has been shown to be indispensable for the replication and cytopathicity of HIV-1 (ref. 41).

Table 1 Quantification of the homologies among retroviral proteins

<i>a</i>	HIV-1	GAG	POL	ENV		F	Central Region			
				EGP	TMP		Q	R	tat	art
	HIV-2	57.7 (95.2)	59.4 (96.6)	39.4 (90.6)	44.8 (91.3)	37.7 (79.7)	34.6 (87.4)	52.2 (85.7)	42.8 (64.6)	44.8 (97)
<i>b</i>		HIV-2		HIV-1		HTLV-I		VISNA		
	HIV-1	59.1 (96.4)		—		ND		ND		
	LAV-Eli	61.6 (96.1)		94 (98.7)		ND		ND		
	LAV-Mal	59 (95.2)		92 (98.7)		ND		ND		
	EIAV	43.8 (92)		41.9 (91.5)		ND		46.7 (90.8)		
	VISNA	43.7 (88.7)		42.2 (94)		ND		—		
	HTLV-I	34.8 (70.5)		33.3 (70.3)		—		ND		
	HTLV-II	ND		ND		62.8 (99.5)		ND		
	BLV	ND		ND		49.5 (93.2)		ND		
	RSV	35.9 (72.3)		34.5 (76.2)		38.2 (86.4)		ND		

The reference protein of each alignment is that listed at the top of the column. Proteins were aligned using the NUCALN program⁶¹ with following parameters: K-tuple 1, window 20, gap penalty 1. Two results are indicated in each case: the amino-acid identity (%) in the aligned domains (that is, excluding the regions of insertion/deletion), and between parentheses the percentage of the length of reference protein that could be aligned. **a**, Homologies between HIV-1 and HIV-2 proteins. For *env*, the calculation was done for the external glycoprotein (EGP, including the signal peptide, whose length is not exactly known in HIV-2), and the transmembrane protein (TMP). **b**, Comparison of the *pol*-encoded proteins of different retroviruses. LAV-Mal and LAV-Eli are Zairian isolates of HIV-1 (ref. 12); EIAV: equine infectious anaemia virus (sequence communicated by Dr S. Aaronson), and visna virus²⁷ are animal lentiviruses; HTLV-I, HTLV-II, BLV^{62–64}, related leukaemogenic retroviruses; RSV) Rous sarcoma virus⁶⁵. ND, not determined.

ENV

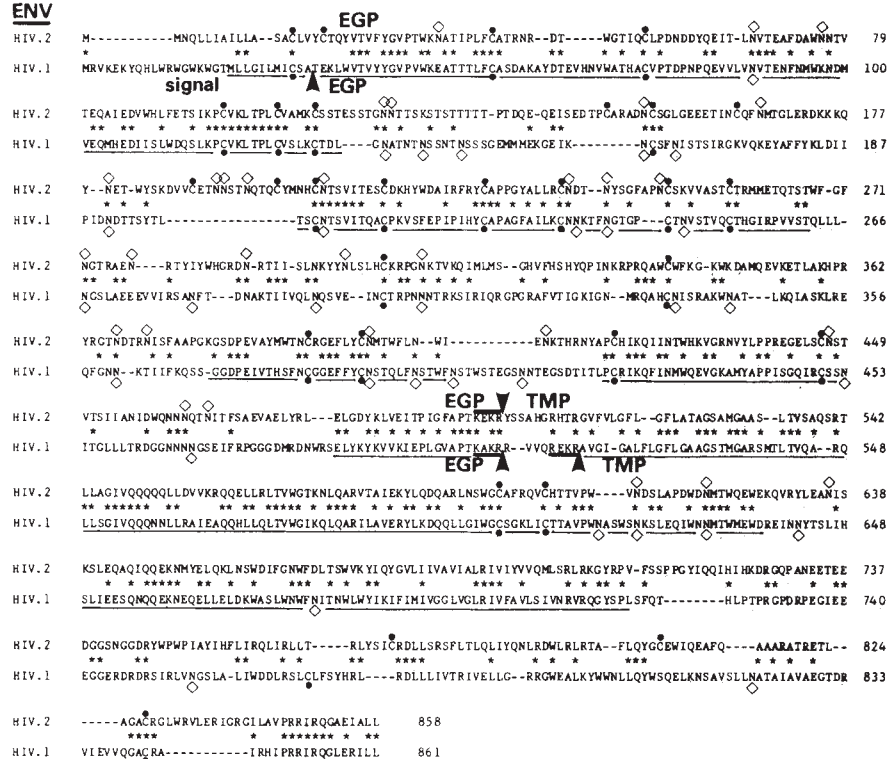
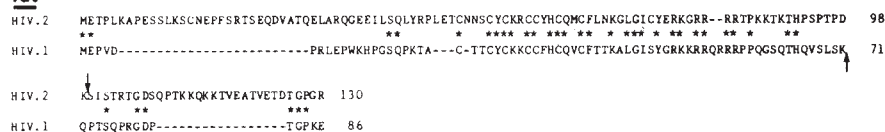
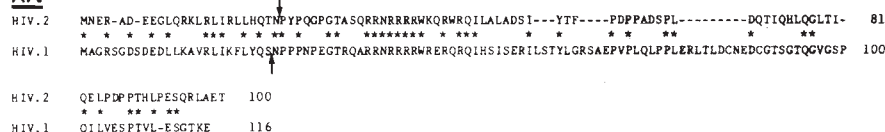


Fig. 4 Alignments of the HIV-1 (BRU isolate, ref. 15) and HIV-2 proteins. Asterisks indicate amino-acid identities. Gaps were introduced to optimize the alignments. In the envelopes, the potential cleavage sites are shown by arrows. EGP, external glycoprotein; TMP, transmembrane protein. ◇, Potential *N*-glycosylation sites; ●, cysteines. The domains of the EGP of HIV-1 that were found to be well-conserved among isolates¹² are underlined. The parts of *tat* and *art* encoded by each of the two exons are separated by an arrow.

Tat



Art



To examine whether transactivation (a property also shared with the ovine visna lentivirus but not with the related caprine arthritis and encephalitis virus⁴²) exists in HIV-2, we constructed a plasmid, called pHIV2-CAT, containing the bacterial chloramphenicol acetyltransferase (CAT) gene under the control of the U3-R region of HIV-2 (225 bp of U3 and 175 bp of R). To test the transactivation of HIV-2, cells were either infected with HIV-2 or mock-infected, and five days later transfected with either pSVCAT (which contains the CAT gene under control of the SV40 early promoter⁴³) or pHIV2-CAT. At the time of transfection, the cells were not producing virus. Nonetheless, we observed a substantial increase in the amount of CAT expression in extracts of HIV-2-infected versus mock-infected cells that had been transfected with pHIV2-CAT (Fig. 5a). The expression of the SV40 early promoter was not affected by HIV-2 infection.

To determine whether the *tat* gene of HIV-1 could transactivate the LTR of HIV-2 and vice versa, we cotransfected SW480 cells⁴⁴ with subgenomic fragments of HIV-1 or HIV-2 and pHIV2-CAT or a plasmid called pHIV1-CAT, which contains U3-R of HIV-1 (the entire U3 and 70 bp of R) directing transcription of the CAT gene. The plasmid pLET (a gift from Dr S. Wain-Hobson) contains the region of the HIV-1 shown by others to encode the HIV-1 *tat* gene^{28,29}. The plasmid pME214, on the other hand, contains HIV-2 sequences between nucleotides 5,786 and 8,571 (Fig. 2), and in particular contains the open reading frames of HIV-2 that share homology with the *tat* gene of HIV-1. In both of these plasmids transcription is driven

by the LTR of the respective virus, and the first AUG of the transcript is the first AUG of the putative *tat* gene. It should be noted that both these plasmids also contain the coding potential for the *art* gene.

Although the SV40 early promoter was not affected by either the HIV-1 *tat* nor the HIV-2 *tat* genes, both HIV-1 and HIV-2 LTRs were substantially activated by the HIV-1 *tat* gene (Fig. 5b). This is perhaps surprising in view of the difference in size of the R region of HIV-1 (where the transactivator responsive region (TAR) resides⁴⁵) and HIV-2. However 35 of the 58 bases present in the first stem-and-loop secondary structure of the TAR region of HIV-1 are conserved, and an analogous stem-and-loop structure with the first 77 bases of R can be drawn for HIV-2 (ref. 33).

The HIV-2 LTR is transactivated over 100-fold by pME214 (Fig. 5b). On the other hand, the HIV-1 LTR is not as well transactivated by this plasmid (~5–20 fold, Fig. 5 and other data not shown). Similar results were obtained after transfection of HeLa and HUT 78 cells (data not shown). These experiments indicate that pME214 encodes a functional *tat* gene. In addition, they indicate that the specificity of the HIV-2 *tat* is somewhat different from that of the HIV-1 *tat*. It will be important to determine whether this observation is isolate-specific.

Origin of human immunodeficiency viruses

We have presented here the complete nucleotide sequence of the retrovirus associated with AIDS in West Africa, HIV-2, and tentatively identified the viral proteins either detected in

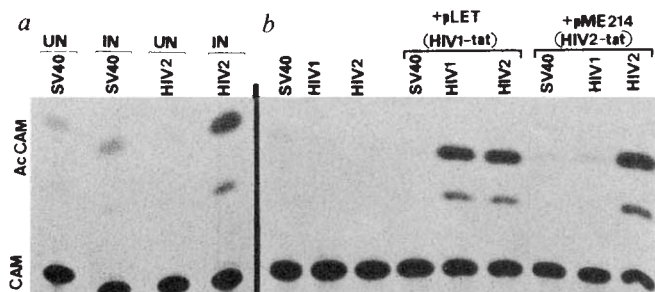


Fig. 5 Transactivation of HIV-2. Chloramphenicol acetyltransferase (CAT) assays were done as described⁵⁹. The unreacted chloramphenicol is marked 'CAM', and the acetylated products are marked 'AcCAM'. All reactions were 1 h with 10% of the cellular extract made 40 h after transfection. The origin of the promoter linked to the CAT gene is indicated above each lane. SV40 indicates the SV40 early promoter, HIV-2 indicates the partial U3 and the entire R sequences of HIV-2 (ROD isolate), and HIV-1 indicates the entire U3 and 70 bp of R of HIV-1 (BRU isolate). **a**, HUT 78 cells were either mock-infected (UN, uninfected) or infected (IN) with HIV-2. Five days post-infection, 3×10^6 cells were transfected with 3 μ g of plasmid in 0.5 ml of Tris-saline without divalent cations for 45 min at 37 °C with 250 μ g ml⁻¹ DEAE-Dextran. **b**, 4×10^5 SW480 cells were cotransfected by the CaCl₂ technique⁶⁰ with 3 μ g of promoter-CAT plasmid and 3 μ g of the indicated plasmid. Salmon sperm DNA was added such that each transfection was 20 μ g ml⁻¹ DNA. This experiment was repeated three times with similar results.

immunoprecipitations with patients' sera, or homologous to proteins previously identified in HIV-1. The two viruses share a similar genomic organization, indicating a common evolutionary origin, but differ significantly in terms of nucleotide and amino-acid sequence: the more-conserved *gag* and *pol* genes respectively display only 56 and 60% nucleotide sequence homology and both less than 60% amino-acid identity. The calculation of the nucleotide sequence homology for the other genes gives even lower values, making HIV-1 and 2 42% homologous overall. This confirms that these two viruses are distinct elements of the HIV family, and cannot be considered as strains of the same virus, according to the recommendations of the international taxonomy committee⁴⁶.

It was previously established that HIV-2 is more related to the simian immunodeficiency viruses (SIV) than it is to HIV-1. The *gag*, *pol* and *env* proteins of SIV and HIV-2 are antigenically cross-reactive, whereas their cross-reactivity to HIV-1 is restricted to some *gag* and *pol* antigens. The amino-terminal amino-acid sequence of the major core protein (corresponding to the p25^{gag} of HIV-1 and p26^{gag} of HIV-2) has been determined in one isolate of SIV obtained from macaques with an AIDS-like disease (MnIV, ref. 26). Out of the 23 amino acids sequenced 21 match with the amino terminus of p26^{gag} of HIV-2, whereas 13 (with one deletion) match to the p25^{gag} of HIV-1. Furthermore, whereas HIV-2 can infect, at least transiently, primate species which are evolutionarily more distantly related to humans (at least baboons and macaques), HIV-1 infects only humans and chimpanzees (R. Desrosiers and P. Fultz, personal communications). In fact, it is not possible from current data to know whether SIV can be classified as distinct from HIV-2 or if they only differ as independent isolates of the same virus.

The almost simultaneous emergence of two foci of AIDS in distinct areas of the African continent is unlikely to be due to the recent emergence of two novel human pathogens, for example by simultaneous trans-species infection by animal retrovirus, or by the mutation of pre-existing non-pathogenic human retroviruses. Indeed, HIV-1 and HIV-2 are obviously retroviruses with a common origin, but they are highly divergent, and it is more likely that their time of divergence is earlier than the beginning of the current epidemics. Therefore a common ancestor, with similar properties and pathogenic potential, prob-

ably existed a long time ago in a human population, and the emergence of the AIDS epidemics is more likely the result of simultaneous modifications of epidemiological parameters in West and Central Africa, such as uncontrolled urbanization, leading to the infection of larger populations.

A question to be addressed is why the HIVs were only recently detected if they existed for a long period. This may be due to the fact that the pathogenicity of an HIV-type retrovirus cannot be revealed until it has spread to a significant portion of the population. First, in areas of Africa with poor medical facilities, where other infections, such as malaria, represent primary causes of morbidity, isolated cases of AIDS could have been an undetectable clinical event. Then, the incubation time can vary considerably, and it cannot still be ruled out that a large fraction of individuals infected by a HIV will remain healthy carriers. In Kenya, HIV-1 seropositivity was first reported in a high fraction of subjects at risk of AIDS (female prostitutes) who were apparently healthy; later, the virus diffused to a larger part of the population, and cases of AIDS were observed⁴⁷. A similar situation could explain the apparent lack of pathogenicity of the retrovirus designated HTLV-IV, but indistinguishable from HIV-2 and SIV by the antigenicity of its proteins^{19,48,49}. The presence of HTLV-IV was identified only in apparently healthy individuals in West Africa, an area where we have observed several typical AIDS cases caused by HIV-2. It is possible that the apparent non-pathogenicity of HTLV-4 is due to a recent epidemic diffusion of HIV-2/HTLV-IV in the West Africa, where AIDS cases still represent a minor fraction of the infected and seropositive individuals, whereas HIV-1 has diffused in major cities of central Africa or the USA some time before.

Implications for vaccines and diagnostics

The risk that HIV-2-infected blood samples may not be detected by standard screens, currently based on the detection of anti-HIV-1 antibodies, makes it important that a way of diagnosing HIV-2 infection is found. As the envelope, and especially its transmembrane part, represents the primary target of the host antibody response to the HIV infection (see ref. 1), antigens from the envelope of HIV-2 will significantly improve the spectrum of the screening tests, allowing the detection of samples infected by HIV-2, and perhaps by other as yet uncharacterized members of the HIV family.

As it shares most of the structural characteristics and biological properties of HIV-1, but displays significant genetic divergence, HIV-2 is a powerful tool in the study of the molecular biology of this group of retroviruses. Among the crucial biological properties common to both HIVs are tropism for CD4-positive cells, and mechanisms of positive regulation of viral expression encoded by viral transactivating factors. We observed that the *tat* of HIV-1 activates the transactivation responsive (TAR) sequences as efficiently in both types of HIV, whereas the *tat* gene of HIV-2 is more efficient on the TAR elements of HIV-2. The *tat* proteins of HIV-1 and 2 have only short homologous sequences, and this will ease the dissection of their function by mutagenesis or using chemically synthesized peptides.

HIV-1 and probably HIV-2 recognize the CD4 surface molecule as a receptor on helper/inducer T lymphocytes and perhaps on other cells expressing the CD4 protein⁵⁰⁻⁵³. In HIV-1, this interaction is mediated by the external envelope glycoprotein (EGP; ref. 52), and an important problem is which of the domain(s) of this protein are involved in that interaction. Indeed, blocking this step of the virus life cycle, either by antibodies or drugs, could be an efficient means for preventing infection or blocking its spread. As the receptor is a constant cellular protein, we can postulate that the binding domain of the envelope is conserved among the CD4-tropic HIVs. The conserved domains of the EGP of HIV-1 and 2 are not numerous, and therefore it becomes possible to demonstrate their possible role in the virus-receptor interaction using a relatively limited

set of site-directed mutations. Given the absence of antigenic cross-reactivity of the envelopes of the two HIVs, this CD4-binding domain is probably not, or only poorly, immunogenic—perhaps because of masking by glycosylation, poor exposure on the virion surface, or mimicking of 'self' antigens. Nevertheless, its presentation to the immune system out of context of the virion, that is, as a peptide, might induce a neutralizing antibody response that is not attained, or attained with only a low efficiency, with the complete native envelope from virions or expression systems^{54–56}.

Conclusion

The comparative analysis of HIV-1 and 2 reveals major genetic differences between retroviruses that share many of their biological properties. They both cause AIDS, are cytopathic *in vitro*,

have a tropism for CD4-bearing cells and have elements *trans*-activating the expression of viral genes acting at the LTR level. The evolutionary potential of these viruses is therefore striking, and we must ask whether other HIVs can emerge as long as a favourable epidemiological situation is provided. We must take advantage of the precise delineation of the conserved structures to understand their molecular biology and develop new therapeutic tools, especially immunoprophylactics.

We thank Drs S. Aaronson, F. Brun-Vézinet, R. Desrosiers, P. Fultz, H. Marquardt and M. Muesing for exchange of scientific information; Drs S. Wain-Hobson, M. L. Michel and A. Louise for providing plasmids; Bernard Caudron for help in preparation of the computer printouts; Dr. Lautaro Perez for discussion and encouragement. M.E. is a fellow of the Leukemia Society of America, Inc.

Received 4 March; accepted 19 March 1987.

- Weiss, R. A. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) Vol. II, 405–485 (Cold Spring Harbor Laboratory, New York, 1985).
- Curran, J. W. *et al. Science* **229**, 1352–1357 (1985).
- Quinn, T.-C., Mann, J. M., Curran, J. W. & Piot, P. *Science* **234**, 955–963 (1986).
- Piot, P. *et al. Lancet* **ii**, 65–69 (1984).
- Clumeck, N. *et al. N. Engl. J. Med.* **310**, 492–497 (1984).
- Barré-Sinoussi, F. *et al. Science* **220**, 868–870 (1983).
- Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497–500 (1984).
- Levy, J. A. *et al. Science* **223**, 840–842 (1984).
- Brun-Vézinet, F. *et al. Science* **226**, 453–456 (1984).
- Ellrodt, A. *et al. Lancet* **i**, 1383–1385 (1984).
- Bailey, A. C. *et al. Lancet* **i**, 359–361 (1985).
- Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63–74 (1986).
- Willey, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5038–5042 (1986).
- Starcich, B. R. *et al. Cell* **45**, 637–648 (1986).
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9–17 (1985).
- Sanchez-Pescador, R. *et al. Science* **227**, 484–492 (1985).
- Ratner, L. *et al. Nature* **313**, 227–284 (1985).
- Muesing, M. A. *et al. Nature* **313**, 450–458 (1985).
- Clavel, F. *et al. Science* **233**, 343–346 (1986).
- Brun-Vézinet, F. *et al. Lancet* **i**, 128–132 (1987).
- Clavel, F. *et al. N. Engl. J. Med.* (in the press).
- Clavel, F. *et al. Nature* **324**, 691–695 (1986).
- Daniel, M. D. *et al. Science* **228**, 1199–1204 (1985).
- Kanki, P. J., Alroy, J. & Essex, M. *Science* **230**, 951–954 (1985).
- Benveniste, R. E. *et al. J. Virol.* **60**, 483–490 (1986).
- Murphy-Corb, M. *et al. Nature* **321**, 435–437 (1986).
- Sonigo, P. *et al. Cell* **42**, 369–382 (1985).
- Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69–73 (1985).
- Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. *Science* **229**, 74–77 (1985).
- Sodroski, J. *et al. Nature* **321**, 412–417 (1986).
- Feinberg, M. B., Jarret, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807–817 (1986).
- Jones, K., Kadonaga, J., Luciw, P. & Tjian, R. *Science* **232**, 755–759 (1986).

- Muesing, M., Smith, D. H. & Capon, D. J. *Cell* **48**, 691–701 (1987).
- Rabson, A. B. *et al. Science* **229**, 1388–1390 (1985).
- Lightfoot, M. *et al. J. Virol.* **60**, 771–775 (1986).
- Allan, J. S. *et al. Science* **228**, 1091–1094 (1985).
- Lee, T. H. *et al. Science* **231**, 1546–1549 (1986).
- Kan, N. C. *et al. Science* **231**, 1553–1555 (1986).
- Allan, J. S. *et al. Science* **230**, 810–813 (1985).
- Miller, J., Mac Laehlan, A. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
- Dayton, A., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941–947 (1986).
- Hess, J. L., Pyper, J. M. & Clements, J. E. *J. Virol.* **60**, 385–393 (1986).
- Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6777–6781 (1982).
- Adachi, A. *et al. J. Virol.* **59**, 284–291 (1986).
- Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813–823 (1985).
- Coffin, J. M. *et al. Nature* **321**, 10 (1986).
- Kreiss, J. K. *et al. N. Engl. J. Med.* **314**, 414–418 (1986).
- Kanki, P. J. *et al. Science* **232**, 238–243 (1986).
- Barin, F. *et al. Lancet* **ii**, 1387–1389 (1985).
- Dalglish, A. G. *et al. Nature* **312**, 763–767 (1984).
- Klatzman, D. *et al. Nature* **312**, 767–768 (1984).
- McDougal, J. S. *et al. Science* **231**, 382–385 (1985).
- Maddon, P. J. *et al. Cell* **47**, 333–348 (1986).
- Weiss, R. A. *et al. Nature* **316**, 69–72 (1985).
- Clavel, F., Klatzman, D. & Montagnier, L. *Lancet* **i**, 879–880 (1985).
- Lasky, L. A. *et al. Science* **223**, 209–212 (1985).
- Messing, J. & Viera, J. *Gene* **19**, 269–276 (1982).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. Cell. Biol.* **2**, 1044–1051 (1982).
- Graham, F. L. & Van der Eb, A. J. *Virology* **52**, 456–467 (1973).
- Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618–3622 (1983).
- Kiyokawa, T., Yoshikura, H., Hattori, S., Seiki, M. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6202–6206 (1984).
- Sagata, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 677–681 (1985).
- Schwartz, D. E., Tizard, R. & Gilbert, W. *Cell* **32**, 853–869 (1983).

LETTERS TO NATURE

Switching phenomena in a new 90-K superconductor

P. H. Hor*, R. L. Meng*, C. W. Chu*, M. K. Wu†, E. Zirngiebl‡, J. D. Thompson‡ & C. Y. Huang§

* Department of Physics, University of Houston, Houston, Texas 77004, USA

† Department of Physics, University of Alabama, Huntsville, Alabama 35899, USA

‡ Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

§ Lockheed Palo Alto Research Laboratory, Palo Alto, California 94304-1191, USA

Recently, Wu *et al.*¹ and Hor *et al.*² have shown that $Y_{1.2}Ba_{0.8}CuO_{4-\delta}$ is a superconductor with a superconducting onset temperature at ~ 92 K as determined by their resistivity and a.c. susceptibility measurements. Because the magnetic properties are important in describing the nature of superconductivity, we have measured the d.c. magnetic moment of this material. Here we show that this material cooled in zero field or in a high field ($H_{cool} >$

90 G) is diamagnetic below $T_{cm} \approx 90$ K, consistent with the previous measurements^{1,2}. However, when the sample is cooled in a small field (≤ 85 G), the magnetization, M , first becomes negative (diamagnetic) below T_{cm} , but further cooling results in a jump of M to a positive value at low temperature. We have also observed this switching by the application of an additional small field when the sample was cooled in a small field.

The $Y_{1.2}Ba_{0.8}CuO_{4-\delta}$ sample was prepared as described in ref. 1. The X-ray diffractograms reveal that the sample has multiple phases, devoid of the K_2NiF_4 structure. From the electrical resistance measurement, the superconducting onset temperature is $T_{co} \approx 94.5$ K and the resistance becomes 'zero' below $T_0 = 92$ K indicating that the sample is a superconductor with a rather narrow transition width. A Quantum Design superconducting quantum interference device (SQUID) magnetometer has been employed to measure the magnetization of the sample as a function of temperature and magnetic field. When the sample is cooled under zero field conditions, we have found that M is diamagnetic below T_{cm} and the susceptibility below ~ 25 K reaches $\sim 35\%$ of that of perfect diamagnetism ($-1/4\pi$).

We have also measured M when the sample is cooled in a field, H_{cool} . In Fig. 1, the magnetization obtained at various

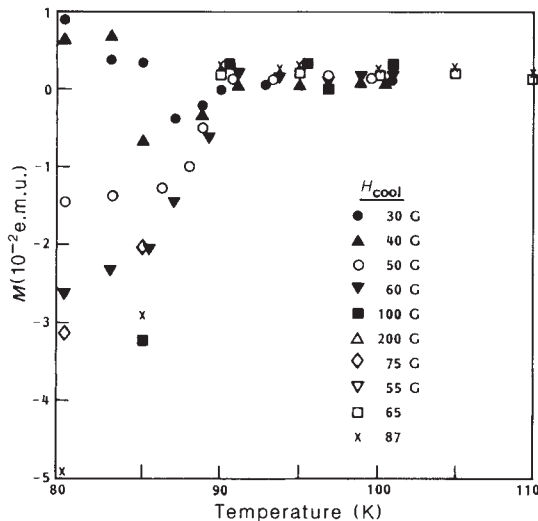


Fig. 1 Magnetization, M , against temperature for $\text{Y}_{1.2}\text{Ba}_{0.8}\text{CuO}_{4-\delta}$ cooled in the magnetic fields, H_{cool} , as shown.

H_{cool} values are displayed. Clearly, M is diamagnetic right below T_{cm} at all H_{cool} values used. The 30-G data (denoted by solid circles) show a jump from a negative value at 87 K to a positive value at 85 K and M increases with decreasing temperature. For $H_{\text{cool}} = 40$ G, the switching takes place at 83 K. As H_{cool} increases, the switching temperature, T_s , decreases (Fig. 2). In the inset, the dependence of the switching temperature on H_{cool} is shown. For $H_{\text{cool}} > 87$ G, the switching disappears. Apart from the switching, the data for warming are identical to those for cooling within experimental error. As shown in Fig. 2, $|M|$ at low temperature is much greater than the paramagnetic magnetization above T_{cm} .

In order to investigate the switching, we have cooled the sample in the magnetic field H_{cool} and then increased the field. As shown in Fig. 3, when the sample is cooled to 5 K in $H_{\text{cool}} = 30$ G, M is positive (open squares). However, as H is increased to 40 G, M switches to a negative value and becomes more negative as H increases. Similar results have been found for $H_{\text{cool}} = 30$ G, $T = 20$ K (open circles); $H_{\text{cool}} = 40$ G, $T = 5$ and 35 K, and $H_{\text{cool}} = 50$ G, $T = 5, 35$ K. However, for $H_{\text{cool}} = 100$ G, $T = 35$ K, M is always negative (solid triangles). For comparison, the zero-field cooled data (denoted by shaded squares) are also displayed. They are more negative than those when the sample is field-cooled. In the inset, we demonstrate the details of the switching as the sample is cooled to 20 K at $H_{\text{cool}} = 40$ G, and then an increment field ΔH is applied. As shown, the switching from $M > 0$ to $M < 0$ takes place at ΔH between 4.0 and 4.5 G. Similar switching phenomena have been observed at different H_{cool} and different temperatures. We would like to point out that the switchings observed are unique to $\text{Y}_{1.2}\text{Ba}_{0.8}\text{CuO}_{4-\delta}$, because we have not observed any switching phenomena for other ceramic superconductors, $(\text{La}_{1-x}\text{Ba}_x)_2\text{CuO}_{4-\delta}$ and $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_{4-\delta}$ using the same magnetometer.

In a recent theoretical paper, Ebner and Stroud³ have investigated the diamagnetic response of weakly linked superconducting clusters. They have shown that the magnetization could be positive or negative depending on H . Our sintered powder sample might consist of weakly linked superconducting clusters; hence, the magnetization could be positive. In addition, in principle, as pointed out by these authors, a large frustrated cluster may choose among numerous competing low nearly degenerate states as in the case of a spin glass⁴, and the switching might be the result of the cluster hopping from one configuration to another. To understand the unexpected switching phenomena

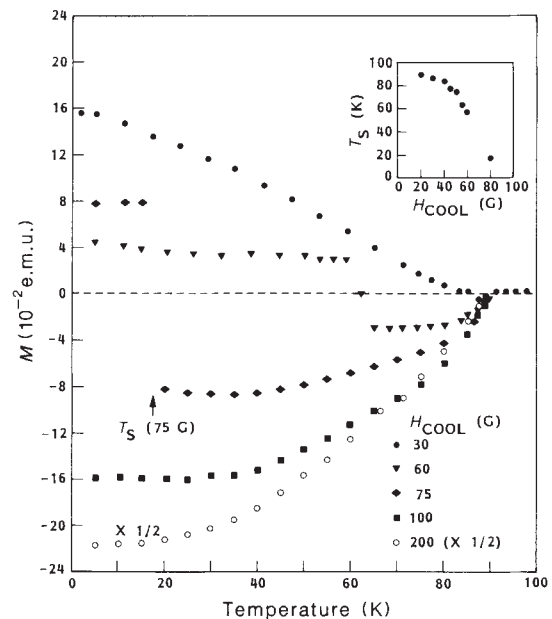


Fig. 2 Magnetization, M , against temperature for $\text{Y}_{1.2}\text{Ba}_{0.8}\text{CuO}_{4-\delta}$ cooled in various magnetic fields H_{cool} . The inset shows the dependence of the switching temperature, T_s , on H_{cool} .

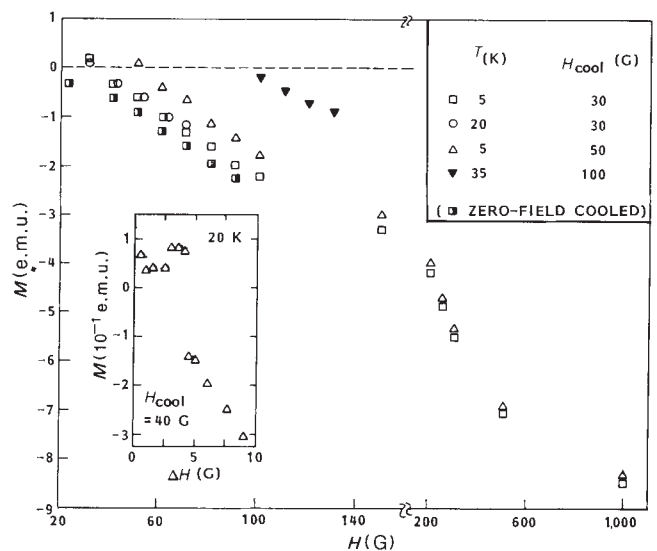


Fig. 3 Magnetization against magnetic field. The $\text{Y}_{1.2}\text{Ba}_{0.8}\text{CuO}_{4-\delta}$ sample was first cooled in H_{cool} , then M is measured as H is increased from H_{cool} . As shown for $H_{\text{cool}} = 30$ and 50 G, M is positive, but M switches to a negative value as H is increased from H_{cool} . In the inset, the switching is shown for the sample cooled to 20 K in 40 G, and then M is measured as the field increase, ΔH , occurs.

more experimental, as well as theoretical work is needed.

We thank Professors C. Lobb and D. Stroud for their stimulating discussions. The work at the University of Houston and the University of Alabama at Huntsville was supported by the NSF and at Los Alamos by the US Department of Energy.

Received 18 March; accepted 25 March 1987.

1. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908-910 (1987).
2. Hor, P. H. *et al. Phys. Rev. Lett.* **58**, 911-912 (1987).
3. Ebner, C. & Stroud, D. *Phys. Rev. B* **31**, 165-171 (1985).
4. Huang, C. Y. *J. Mag. Mag. Mater.* **51**, 1-74 (1985).

Analysis of a proposed crucial test of quantum mechanics

M. J. Collett & R. Loudon

Physics Department, Essex University, Colchester CO4 3SQ, UK

An experiment based on an extension of the Einstein-Podolsky-Rosen argument has been proposed by Popper¹ as a crucial test of the Copenhagen interpretation of quantum mechanics. Here we show, by a slightly more complete version of Popper's analysis, although still at a relatively primitive level of sophistication, that the proposed experiment does not in fact provide such a test.

The form of the experiment as outlined by Popper is represented in Fig. 1. A source emits pairs of particles in opposite directions, as for example in the decay of positronium into γ -rays. It is assumed that the rate of emission is sufficiently low that at most one pair of particles is in flight at any time. Some of the particles pass through a slit at R on the right-hand side of the source, giving us knowledge of their position in the y -direction. The presence of the slit produces an uncertainty-principle scatter in the y -momenta of the particles so that they fall on a range of the detectors arrayed behind the slit. On the left-hand side of the source, at L, any slit is so wide that we can ignore it, but there is an array of detectors wired to operate in coincidence, so that only particles paired with those detected through the slit R are registered. Because the pairs of particles have opposite momenta, we have thus indirectly measured the y -coordinates of the left-hand particles as they pass the point L.

According to the argument of Popper, the standard interpretation of quantum mechanics predicts that the indirect measurement ('mere knowledge') of the y -coordinates of the particles at L should produce an increased scatter on the left as the slit

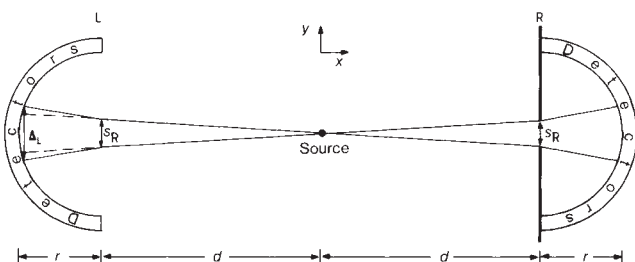


Fig. 1 Experiment proposed by Popper.

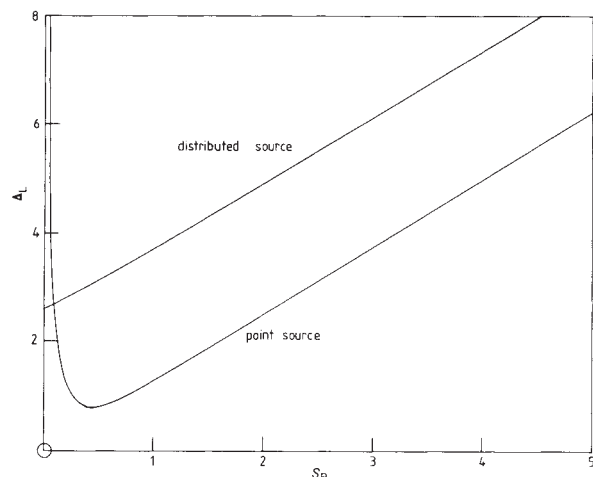


Fig. 2 Variation of the range Δ_L of activated detectors with slit width s_R , taken from equation (3) for the point source and equation (9) for the distributed source, with $r/d = \frac{1}{4}$ and $\Delta y = 1$.

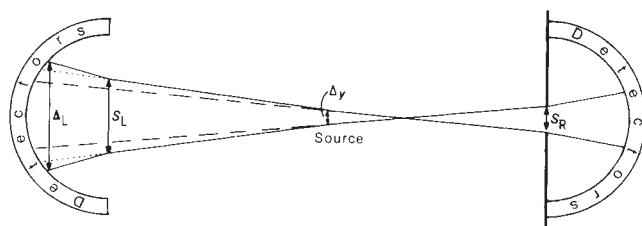


Fig. 3 Modified experimental arrangement with allowance for source uncertainties.

R is made narrower, just as if there were a slit at L of the same width as the slit at R. If, on the other hand, mere knowledge of the y -coordinates of the particles at L does not affect their behaviour, we should expect that, as the slit R is made narrower, although the detectors on the right would register an increased spread, those on the left would give a reduced one, as represented by the dashed lines in Fig. 1.

More quantitatively, the range of activated detectors on the left, according to the standard interpretation, is given by

$$\Delta_L^2 = \left(\frac{d+r}{d} s_R \right)^2 + \left(\frac{r\lambda}{4\pi s_R} \right)^2 \quad (1)$$

where λ is the particle wavelength and the notation for the various lengths is as shown in Fig. 1. The first term on the right-hand side of equation (1) is the geometrical contribution of the y -coordinate uncertainty and the second term is that of the y -momentum uncertainty produced by the virtual slit at L. It is convenient to re-express equation (1) in dimensionless form by measuring all lengths in units of

$$(d\lambda/4\pi)^{1/2} \quad (2)$$

whence,

$$\Delta_L^2 = \left(1 + \frac{r}{d} \right)^2 s_R^2 + \left(\frac{r}{d} \right)^2 \frac{1}{s_R^2} \quad (3)$$

The form of variation of Δ_L with slit width s_R is shown by the curve labelled 'point source' in Fig. 2. For

$$s_R < [r/(d+r)]^{1/2} \quad (4)$$

the final term in equation (1) dominates to give a larger Δ_L for a smaller s_R . Without the 'mere knowledge' effect, only the first term on the right of (3) appears and Δ_L is proportional to s_R .

There is, however, a major flaw in the above argument: it assumes that the positronium, or other source, is stationary at a fixed position. A more careful analysis must allow for uncertainty-principle limitations on the accuracy with which the state of the source can be specified. This state of affairs is represented in Fig. 3, and there are now three contributions to the spread of activated detectors on the left. First, there is a geometrical contribution. With allowance for the finite uncertainty Δy in source position, consideration of the geometry of the dashed lines in Fig. 3 gives a contribution

$$\Delta_L' = 2\Delta y + s_R + (r/d)(\Delta y + s_R) \quad (5)$$

Second, with the source in a minimum uncertainty state, its momentum spread $\hbar/2\Delta y$ produces a corresponding relaxation of the requirement that the pairs of particles are emitted with exactly equal and opposite momenta. The resulting uncertainty at the left-hand detectors, represented by the dotted lines in Fig. 3, is

$$\Delta_L'' = (d+r)\lambda/4\pi\Delta y \quad (6)$$

Third, the virtual slit of width s_R at L produces a contribution

$$\Delta_L''' = r\lambda/4\pi s_L \quad (7)$$

We take the size of s_L to be that deduced from the experimenter's knowledge of the source and right-hand slit dimensions, and

there are thus two contributions analogous to the first two mentioned above, with a total width given by

$$s_L^2 = (2\Delta y + s_R)^2 + (d\lambda/4\pi\Delta y)^2 \quad (8)$$

The total spread of activated detectors given by a combination of the three contributions, with all lengths expressed in the units of (2), is given by the dimensionless expression

$$\Delta_L^2 = \{2\Delta y + s_R + \frac{r}{d}(\Delta y + s_R)\}^2 + \frac{(d+r)^2}{d^2(\Delta y)^2} + \frac{(r/d)^2}{(2\Delta y + s_R)^2 + (\Delta y)^2} \quad (9)$$

This reduces to equation (3) if the terms that involve the source position and momentum uncertainties are removed. However, if these terms are retained, the variation of Δ_L with s_R predicted by equation (9) is quite different from that predicted by equation (3). It is not difficult to show by differentiation that Δ_L is everywhere an increasing function of s_R . The effect of the source uncertainties is to make it impossible for the virtual slit ever to be sufficiently narrow for its contribution to dominate that of the geometric term. There is no choice of experimental parameters for which a reduction in the width of the slit causes an increase in the spread of particles on the opposite side of the source. An example of the monotonic, almost linear increase of Δ_L with s_R is provided by the curve labelled 'distributed source' in Fig. 2.

The uncertainties in transverse momenta produced by the finite slit and source dimensions are of course similar to those obtained from classical diffraction theory, and the above results for Δ_L do not depend on Planck's constant. Clauser² has pre-

viously shown that the quantum and classical theories produce the same angular spread in the correlation of pairs of γ -rays emitted in positronium decay. Very similar considerations apply to a variant of the experiment described above in which spatial correlations are sought between pairs of photons generated in parametric splitting in a nonlinear optical crystal^{3,4}. The above results apply to this experiment, with Δy now representing the width of the primary light beam.

It has been shown that source uncertainty effects in the experiment proposed by Popper¹ remove the distinctive increase in left-hand beam divergence with reduction in right-hand slit width that he ascribed to the Copenhagen interpretation of quantum mechanics. Both these lengths are indeed expected to decrease together, and the experiment cannot provide a crucial test of quantum-mechanical interpretations. Finally, it should be mentioned that Popper's use of uncertainty relations has previously been criticized by Sudbery⁵, but without the more detailed derivations presented here, and his more qualitative conclusions on the effects of the slit on the coincident activations of the left-hand detectors do not agree with ours.

We thank E. R. Pike for drawing our attention to Popper's proposed experiment and for helpful discussions. M.J.C. thanks the Association of Commonwealth Universities for support in the form of a Commonwealth Scholarship.

Received 8 December 1986; accepted 24 February 1987.

1. Popper, K. R. *Quantum Theory and the Schism in Physics* 27-29 (Hutchinson, London, 1982).
2. Clauser, J. F. *Coherence and Quantum Optics* (eds Mandel, L. & Wolf, E.) 815-827 (Plenum, New York, 1973).
3. Burnham, D. C. & Weinberg, D. L. *Phys. Rev. Lett.* **25**, 84-87 (1970).
4. Jakeman, E. & Jefferson, J. H. *Optica Acta* **33**, 557-576 (1986).
5. Sudbery, A. *Phil. Sci.* **52**, 470-476 (1985).

Large-scale peculiar velocities from cosmic strings

E. P. S. Shellard*, R. H. Brandenberger*,
N. Kaiser† & N. Turok‡

* Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge CB3 9EW, UK

† Institute of Astronomy, University of Cambridge,
Cambridge CB3 0HD, UK

‡ Blackett, Laboratory, Imperial College, London SW7 2BZ, UK

Recent observations¹⁻³ indicate that large regions of the Universe $60 h^{-1}$ Mpc in radius (where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$) may be streaming coherently with velocities up to $1,000 \text{ km s}^{-1}$ with respect to the rest frame determined by the cosmic microwave background. In this letter the peculiar velocities which arise in the cosmic string theory of galaxy formation are estimated using numerical simulations. In contrast to models of formation of structure based on primordial linear energy density fluctuations with random phases, the distribution of velocities is highly non-gaussian on scales smaller than $50 h^{-1}$ Mpc. The most likely velocities are approximately scale independent on scales between 5 and $50 h^{-1}$ Mpc.

Large-scale velocities of the above magnitude in conjunction with velocities of 600 km s^{-1} on the scale of the local group ($5 h^{-1}$ Mpc), are in conflict with most of the models of formation of structure based on linear energy density fluctuations with a Zel'dovich spectrum and random phases^{4,5}. Hence, it is of interest to estimate the predicted peculiar streaming velocities in the cosmic string theory of galaxy formation⁶⁻⁸, a model with non-random phases. Nonrandom phases of the initial spectrum of energy density fluctuations can lead to a non-gaussian distribution of velocities. We show, using numerical simulations and under several simplifying assumptions, that this is indeed the

case in the cosmic string model on scales smaller than $50 h^{-1}$ Mpc. If the value of the mass per unit length μ is taken to be

$$G\mu = 2 \times 10^{-6} \quad (1)$$

(where G is Newton's gravitational constant) the r.m.s. velocity on a scale of $60 h^{-1}$ Mpc is about 150 km s^{-1} , comparable to the results in the usual cold dark matter model^{4,5}. The analytical derivation of these results will be discussed in a subsequent paper. Unless otherwise indicated we use units in which the speed of light c is 1.

The peculiar (non-Hubble) velocities considered here are obtained by comparing redshifts of galaxies with their distances. These distances are estimated by assuming that galaxies satisfy a universal relationship between luminosity L and, for example, velocity dispersion σ (ellipticals). Because the determination of the peculiar velocities involves considerable theoretical prejudice there is ample scope for potential systematic errors.

Collins *et al.*¹ consider an all-sky sample of galaxies with a mean redshift of $50 h^{-1}$ Mpc and compute a velocity of $1,000 \pm 300 \text{ km s}^{-1}$, confirming the results of an earlier investigation by Rubin *et al.*⁹ based on the same sample. Burstein *et al.*² considered elliptical galaxies contained in 10 clusters located inside a sphere of radius $60 h^{-1}$ Mpc and detected a net streaming motion of $700 \pm 300 \text{ km s}^{-1}$ in a direction agreeing roughly with the result of Collins *et al.*

On the other hand, Aaronson *et al.*³ considered spiral galaxies in a region of similar depth but restricted to the narrow declination range accessible to the Arecibo telescope. They detected no statistically significant net streaming motion. These observations are not necessarily inconsistent, because the net velocity determined in refs 1 and 2 is mainly orthogonal to the plane accessible to the Arecibo telescope.

In cosmological models peculiar velocities are induced by energy density perturbations. Consider the peculiar velocity $\delta v(\lambda)$ of a galaxy a distance λ from an observer. The Hubble

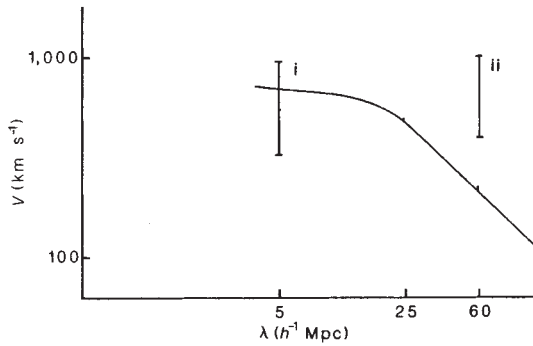


Fig. 1 The observed velocities on the scale of the local group (i) and on large scales (ii). The curve is a sketch of the predicted velocities in the cold dark matter model with random phases.

expansion velocity $v(\lambda)$ is $H\lambda$ where H is the present Hubble constant; $(\delta\rho/\rho)(\lambda)$ shall denote the r.m.s. relative mass perturbation in a sphere of radius λ . Then the r.m.s. velocity $\delta v(\lambda)$ is given in ref. 10.

$$\frac{\delta v}{v}(\lambda) = \frac{1}{3} \frac{\delta\rho}{\rho}(\lambda) \quad (2)$$

The usual cosmological models (in particular the cold and hot dark matter models) assume primordial linear energy density perturbations with random phases and a scale dependence given by

$$\frac{\delta\rho}{\rho}(\lambda) = A \left(\frac{2\pi}{\lambda} \right)^2 \eta^2 \quad (3)$$

where $\eta(t) \sim t^{1/3}$ is conformal time and A is the amplitude. The distribution of velocities will be gaussian, peaked about the value

$$\delta v = \frac{1}{3} H \lambda A \left(\frac{2\pi}{\lambda} \right)^2 \eta^2 \sim \lambda^{-1} \quad (4)$$

Equations (3) and (4) are valid on scales larger than the co-moving Hubble radius at the time t_{eq} of equal matter and radiation which is about $20 h^{-1}$ Mpc. On smaller scales the slope is less, but the exact behaviour is model dependent. The main point is that given the scaling in equation (4) it is extremely unlikely to observe velocities of 600 km s^{-1} and 700 km s^{-1} on scales of $5 h^{-1}$ Mpc and $60 h^{-1}$ Mpc respectively independent of the value of the amplitude A (Fig. 1).

It is therefore interesting to study the predicted peculiar velocities in a cosmological model with nonrandom phases in which the distribution of velocities is expected to be non-gaussian. Such a model is the cosmic string theory of galaxy formation⁶⁻⁸.

In the cosmic string theory¹¹ structures such as galaxies and clusters of galaxies form by accretion onto seed masses. The seeds are loops of physical radius R , length βR (with $\beta \approx 9$; ref. 12) and mass per unit length μ . The number density of seed loops with radius between R and $R+dR$ is given by

$$\begin{aligned} n(R, t) dR &= \nu_m R^{-2} t^{-2} dR & R > t_{eq} \\ &= \nu_r R^{-5/2} t_{eq}^{1/2} t^{-2} dR & R < t_{eq} \end{aligned} \quad (5)$$

ν_m and ν_r are constant coefficients depending on the rate of production of loops in an expanding universe and are of the order $\nu \sim 10^{-2}$ (ref. 12). Soon, numerical simulations will determine ν_m and ν_r more precisely.

The main point is that the mean separation between loops of radius $\sim R$ is much larger than R . Thus the distribution of the mass perturbation is very different from that in the usual models. When expanded into plane wave modes the spectrum of density perturbations has nonrandom phases.

Our analysis of the peculiar velocities in the cosmic string theory is simplified in several respects. We assume that the

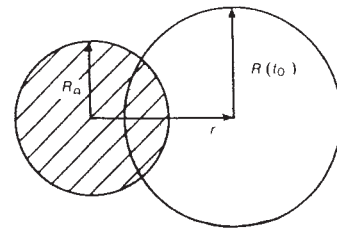


Fig. 2 Sketch of a loop of radius R at a distance r from the observation region V with radius R_0 . $R(t_0)$ is the co-moving distance corresponding to R at the time when the loop of radius R crosses the Hubble radius.

accretion of matter onto a string loop can be described by the spherical collapse model¹⁰. No doubt this is a severe simplification, but it should not effect the order of magnitude of the results, in particular as it can be shown in linear perturbation theory that the total mass accreted onto a loop agrees with that given by the spherical collapse model (E. Bertschinger, preprint, University of California, Berkeley, 1986). This remains true even if the loop is moving and accretion produces an elongated structure.

A second assumption is that the loops are randomly distributed with a number density given by equation (5). Loops are formed when an infinite string crosses itself. These primary loops often fragment into secondary loops. Both effects will contribute to a nonrandom distribution of loops. In future work we intend to apply our analysis to realistic distributions of loops which can be obtained from the numerical simulations of the evolution of the string network by Albrecht and Turok¹².

Finally, in the cosmic string theory matter not only accretes onto loops, but also onto the wakes which form behind moving infinite strings (see refs 13, 14 and A. Stebbins *et al.*, preprint, Cambridge University, 1986). The wakes will therefore also induce peculiar velocities. In this report we neglect this effect, but in a subsequent paper we will return to the issue.

Here we estimate the maximal velocity due to a single string loop as a function of the size of the region over which the average is taken.

R_0 shall denote the radius of the observation region V whose net streaming velocity is to be determined. Consider a loop of radius R at a distance r from the centre of V (Fig. 2). The induced net streaming velocity of V is maximal if $r \sim R_0$. We shall denote the value by $\delta v(R, R_0)$. The next step is to maximize $\delta v(R, R_0)$ by varying R . The maximal value $\delta v(R_0)$ arises if R is equal to R_m , the Hubble radius at the time $t_i(R_0)$ when the co-moving length R_0 enters the Hubble radius, that is

$$R_m = \frac{2}{3} t_i(R_0) = \left(\frac{2}{3} \right)^2 R_0^3 t_0^{-2} \quad (6)$$

where t_0 is the present time. The above analysis is valid on large scales, that is scales for which $t_i(R_0) \geq t_{eq}$. By equation (2) and taking into account that $\delta\rho/\rho$ will only begin to grow at $t_i(R_0)$, we find

$$\delta v(R_0) = \beta G \mu \left(\frac{t_0}{R_0} \right) \sim 90 (G \mu)_6 \text{ km s}^{-1} \quad (7)$$

where $(G \mu)_6$ is $G \mu$ in units 10^{-6} .

Smaller loops with radius $R = f R_m$ (where f is some constant smaller than 1) are less massive. Even though the energy density perturbation can start to grow earlier by a factor of $f^{2/3}$ in redshift (provided $R > \frac{2}{3} t_{eq}$), the final density perturbation is smaller by a factor $f^{1/3}$. Larger loops ($f > 1$) are more massive, but the density perturbation can only start to grow when the co-moving scale $R_0(t)$ corresponding to R_0 exceeds R . This occurs a factor f in redshift later. Hence the velocity perturbation has the same magnitude as that from a loop of radius R_m . However, larger loops are less numerous, their number density being suppressed by $R^{-3/2}$. Hence, contrary to what one might

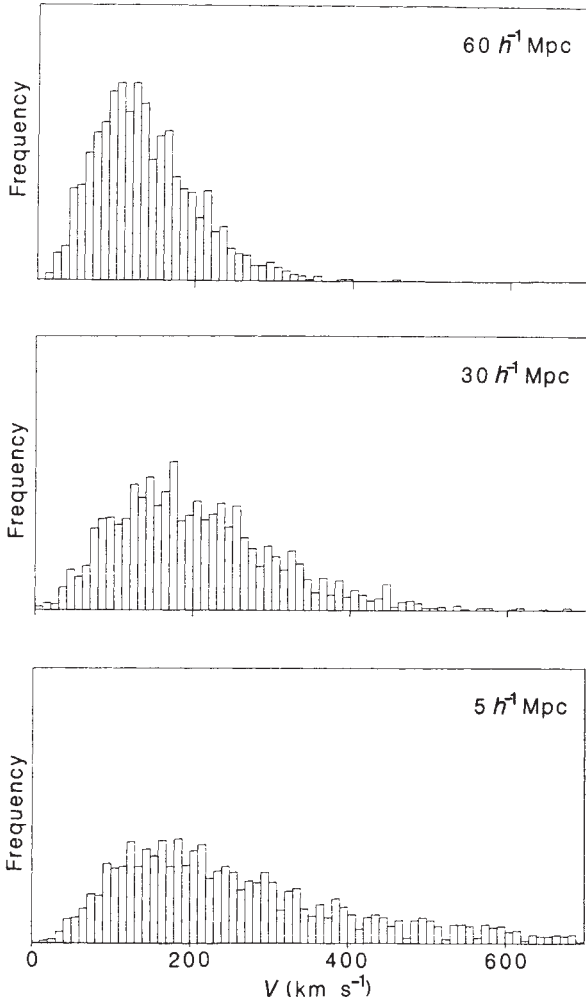


Fig. 3 The distribution of velocities on three different scales in the cosmic string model.

first expect, these larger loops do not provide the maximal effect.

Note that the distribution of velocities from a single loop is bounded above. Larger velocities than those given by equation (7) are not due to larger individual loops, but due to the superposition of the contributions from smaller loops.

To determine the net streaming velocity induced by a distribution of loops numerical simulations were carried out.

In a sphere of radius $300 h^{-1} \text{ Mpc}$ a distribution of loops with number density chosen according to (5) and with randomly chosen centres was laid down. The peculiar velocity $v(p)$ for a point p due to the distribution of loops was calculated by vectorially adding the velocities due to each loop. The net streaming velocity for an observation region V being a sphere with radius R_0 was determined by averaging the point velocities $v(p)$ for all grid points inside V .

The velocities $v(p)$ from a single loop were calculated using the spherical collapse assumption. We need the peculiar velocity $v_R(r)$ of a point located a distance r from a loop with radius R . Consider first large loops formed after t_{eq} ($R > t_{\text{eq}}$) and let z_R denote the redshift at the time the loop entered the Hubble radius. Then for $r = Rz_R$ (the co-moving distance which had the physical magnitude R at redshift z_R)

$$v_R(Rz_R) = \left(\frac{2}{3}\right)^{-2/3} \beta G\mu \left(\frac{t_0}{R}\right)^{1/3} \quad (8)$$

For $v_R(R) < r < Rz_R$ the velocity scales as r^{-1} , because the perturbation on a length scale r can only start to grow when $r(t) > R$ (where $r(t)$ is the physical distance corresponding to the comov-

ing length r); $r_v(R)$ is the virialized radius. In the context of the spherical collapse model $r_v(R)$ is half the radius which is 'turning around' at the present time¹⁰.

$$r_v(R) = \frac{1}{2} t_0 \left(\frac{9}{2} \beta G\mu\right)^{1/2} \quad (9)$$

For $r < r_v(R)$ we set $v_R(r) = 0$. For $Rz_R < r < 2Rz_R$, the upper bound being the distance corresponding to the coming horizon at a redshift z_R , the velocity drops as r^{-2} since the excess relative mass perturbation drops as r^{-3} and the Hubble velocity increases linearly in r . For even larger distances there are two additional potential suppression mechanisms. First, the perturbation in a sphere of radius r can only start to grow when r enters the horizon. This changes the r^{-2} scaling to r^{-4} . Second, there is the issue of compensating underdensities¹⁵. By a relativistic generalization of energy conservation¹⁶⁻¹⁸, the energy density perturbation integrated over space must vanish. The strings constitute an overdensity in scalar field energy. The underdensity may be an underdensity in radiation, in dust or in small loops. This issue will be addressed in detail elsewhere. Here we assume that the underdensity is in radiation. It is thus spread out over the entire horizon volume and does not inhibit the growth of perturbations on scales inside the horizon. If the underdensity were in dust or small loops it would only travel a distance $2Rz_R$ from the overdensity. In this case there would be an additional dipole suppression on larger scales, changing the r^{-4} scaling to a r^{-5} dependence. Numerical simulations with this choice show that the velocities are not changed by more than 25%.

For small loops ($R < \frac{3}{2} t_{\text{eq}}$) the analysis is similar. The virialized radius is always larger than Rz_{eq} , the co-moving scale corresponding to the radius of the loop at the time the perturbation about the loop can start to grow.

$$r_v(R) = \frac{1}{2} \left(\frac{9}{2} \beta G\mu\right)^{1/3} \left(\frac{R}{t_0}\right)^{1/3} t_0 z_{\text{eq}}^{1/3} \quad (10)$$

Hence there is no interval with $v_R(r) \sim r^{-1}$. On length scales r outside the horizon at t_{eq} (that is $r > 3t_{\text{eq}}z_{\text{eq}}$) there is again an extra r^{-2} suppression which turns the r^{-2} dependence of $v_R(r)$ into a r^{-4} dependence. For $r_v < r < 3t_{\text{eq}}z_{\text{eq}}$

$$v_R(r) = \frac{3}{2} \beta G\mu \left(\frac{t_0}{r}\right)^2 \quad (11)$$

In Table 1 we summarize the velocity profile about large and small loops.

There are uncertainties hidden in the above model which will affect the final magnitude of the peculiar velocities. One important factor is the precise distance at which the dipole suppression starts, another, the point at which the r^{-4} scaling begins. Our numerical analysis shows that different choices do change the overall magnitude of $\delta v(R_0)$ but do not affect the scale dependence or the shape of the distribution.

Our numerical results are summarized in Figs 3a-c and 4. We used $h = \frac{1}{2}$, $\Omega = 1$ (Ω is the ratio between energy density and critical energy density for a flat universe), $G\mu = 2 \times 10^{-6}$, $\nu = 10^{-2}$ and $\beta = 9$. The lower cutoff for loop radius was $10^{-1} R_0(t_{\text{eq}})$. By modifying this cutoff we checked that the results are fairly independent of it. The justification for this cutoff comes from analytical results which show that the r.m.s. velocities obtain their dominant contribution from loops of radius $R_0(t_{\text{eq}})$ (R.H.B. *et al.*, preprint). In particular the contribution from loops with a radius smaller than $R < R_0(t_{\text{eq}})$ scales as $R^{1/2}$.

Table 1 Velocity profile about large and small loops

$R > \frac{3}{2} t_{\text{eq}}$		$R < \frac{3}{2} t_{\text{eq}}$	
$r_v(R) < r < Rz_R$:	$v(r) \sim r^{-1}$	$r_v(R) < r < 2Rz_R$:	$v(r) \sim r^{-2}$
$Rz_R < r < 2Rz_R$:	$v(r) \sim r^{-2}$	$2Rz_R < r < 3t_{\text{eq}}z_{\text{eq}}$:	$v(r) \sim r^{-2}$
$2Rz_R < r$:	$v(r) \sim r^{-4}$	$3t_{\text{eq}}z_{\text{eq}} < r$:	$v(r) \sim r^{-4}$

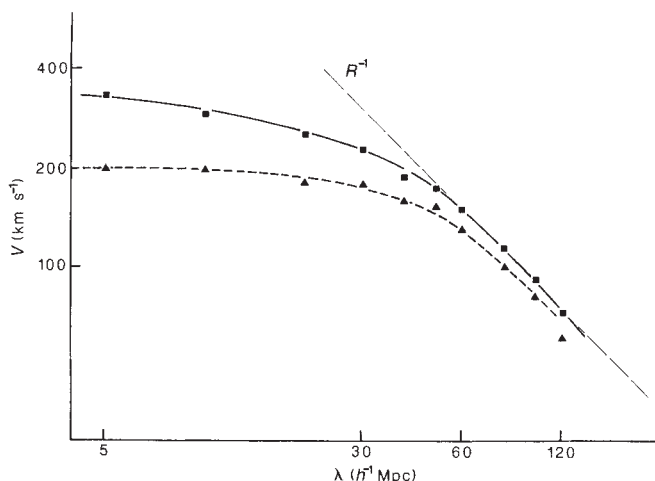


Fig. 4 The r.m.s. velocities (squares) and the most likely velocities (triangles) as a function of radius of the observation region in the cosmic string model.

Figure 3a-c shows the distribution of velocities for $60 h^{-1}$ Mpc, $30 h^{-1}$ Mpc and $5 h^{-1}$ Mpc. Figure 4 shows the scale dependence of the most likely r.m.s. and maximal velocities. The most likely velocity is given by the peak in the velocity distribution. For the three scales of Fig. 3 the sample space exceeds 2,000.

The most remarkable feature in Fig. 3 is the non-gaussian velocity distribution on small scales. The r.m.s. velocity at $5 h^{-1}$ Mpc exceeds the most likely velocity (defined by the peak in the velocity distribution) by a factor of almost 2. On scales larger than $50 h^{-1}$ Mpc the difference between r.m.s. and most likely velocities becomes statistically insignificant.

From Fig. 4 we can see that the most likely velocity is almost constant on scales smaller than $50 h^{-1}$ Mpc. In a subsequent paper we will give a theoretical explanation of this feature.

The above two points clearly distinguish the cosmic string model, a model with nonrandom phases, from the usual cosmological models. The scale independence of the most likely velocity makes it, in principle, easier to reconcile the recent observations of velocities on a scale of $60 h^{-1}$ Mpc with the local group velocity.

For the value $G\mu = 2 \times 10^{-6}$ favoured from previous calculations (with cold dark matter, $h = \frac{1}{2}$ and $\Omega = 1$)¹⁹ the r.m.s. velocities are comparable with those in the usual cold dark matter model. The maximal velocity of 460 km s^{-1} for $R_0 = 60 h^{-1}$ Mpc is still only marginally consistent with the results of ref. 2. A value $G\mu = 5 \times 10^{-6}$ would fit the data better. A slightly larger value of $G\mu$ may arise in the cosmic string theory if the dark matter is hot.

The uncertainty in the value of $G\mu$ is just one of many uncertainties in the analysis. Our model is based on several idealizations: the loops will not be uniformly distributed, β is not fixed but is a random variable and wakes have been entirely neglected. The amplitude is also sensitive to the value of ν . Recent simulations by one of us (N.T.) have shown that ν is larger by a factor of at least 3 (0.03) for loops produced in the matter dominated era. To assess the implications of a larger value of ν we have run simulations in which ν was taken to be 0.03 and 0.1 for loops created in the matter-dominated era. We found that the r.m.s. velocities on large scales increase and that the spectrum of most likely velocities remains approximately flat even to a larger scale. On a scale of $60 h^{-1}$ Mpc the r.m.s. velocities were found to be 225 km s^{-1} for $\nu = 0.03$ and 375 km s^{-1} for $\nu = 0.1$ which corresponds to an increase over the value for $\nu = 0.01$ by a factor of 1.5 and 2.5 respectively.

Despite these uncertainties which are currently inherent in the cosmic string model, we expect our analysis to give the

correct scaling behaviour and the right order of magnitude. These simulations will be discussed elsewhere.

We are grateful to George Efstathiou and Martin Rees for helpful discussions.

Received 25 November 1986; accepted 9 February 1987.

1. Collins, C., Joseph, R. & Robertson, N. *Nature* **320**, 506-508 (1986).
2. Burstein, D. *et al.* in *The Extragalactic Distance Scale and Deviations from the Hubble Flow* (ed. Madore, B.) (Reidel, Dordrecht 1986).
3. Aaronson, M. *et al.* *Astrophys. J.* **302**, 536-563 (1986).
4. Bond, J. in *The Extragalactic Distance Scale and Deviations from the Hubble Flow* (ed. Madore, B.) (Reidel, Dordrecht 1986).
5. Vittorio, N., Juszkiewicz, R. & Davis, M. *Nature* **323**, 132-133 (1986).
6. Zel'dovich, Ya. B. *Mon. Not. R. astr. Soc.* **192**, 663-667 (1980).
7. Vilenkin, A. *Phys. Rev. Lett.* **46**, 1169-1173 (1981).
8. Kibble, T. W. B. *Phys. Rep.* **67**, 183-199 (1980).
9. Rubin, V., Ford, W., Thonnard, N., Roberts, M. & Graham, J. *Astr. J.* **81**, 687-718 (1976).
10. Peebles, P. J. E. *The Large-Scale Structure of the Universe* (Princeton University Press, 1980).
11. Vilenkin, A. *Phys. Rep.* **121**, 263-315 (1985).
12. Albrecht, A. & Turok, N. *Phys. Rev. Lett.* **54**, 1868-1871 (1985).
13. Silk, J. & Vilenkin, A. *Phys. Rev. Lett.* **53**, 1700-1703 (1984).
14. Vachaspati, T. *Phys. Rev. Lett.* **57**, 1655-1657 (1986).
15. Traschen, J., Turok, N. & Brandenberger, R. *Phys. Rev. D* **34**, 919-930 (1986).
16. Zel'dovich, Ya. B. *Adv. Astr. Astrophys.* **3**, 241-379 (1965).
17. Hogan, C., Kaiser, N. & Rees, M. *Phil. Trans. R. Soc. A* **307**, 97-110 (1982).
18. Traschen, J. *Phys. Rev. D* **29**, 1563-1574 (1984).
19. Turok, N. & Brandenberger, R. *Phys. Rev. D* **33**, 2175-2181 (1986).

Extreme scattering events caused by compact structures in the interstellar medium

R. L. Fiedler*, B. Dennison*†, K. J. Johnston* & A. Hewish‡

* E. O. Hulburt Center for Space Research, Naval Research Laboratory, Washington DC 20375, USA

† Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, 24061, USA

‡ Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Daily flux density measurements of 36 extragalactic radio sources over a seven year period reveal several unusual minima in the light curves that do not follow typical source variations. The observations were performed using the Green Bank interferometer operating at 2.7 and 8.1 GHz¹. The most significant departure from typical source variability occurred at both frequencies in the quasar 0954 + 658 between 1980.95 and 1981.3. It is unlikely that this event can be explained by intrinsic variability. The most satisfactory explanation appears to involve refractive focusing by small-scale inhomogeneities in an ionized structure in the interstellar medium.

Figure 1 shows the flux density variation of the quasar 0954 + 658 at both frequencies during the period 1980.5-1981.7. The flux densities are believed to be accurate to 4%. Calibration and data reduction procedures are discussed by Fiedler *et al.*¹. As a check against possible instrumental effects the parallel polarization correlators were checked for all six baselines of the interferometer. We found that all 12 correlators at each frequency showed the same variations.

Explanations involving intrinsic variations seem unlikely to us. In particular, the strong spikes at 8.1 GHz and the relatively constant flux minimum at 2.7 GHz are difficult to explain. Furthermore, the typical variations of 0954 + 658 bear no relation whatsoever to this event. That the unusual variations between 1980.95 and 1981.3 at 2.7 and 8.1 GHz appear centred on one another argues against the intrinsic variability models that rely on an evolving emitter, where the lower frequency flux density variations lag those at higher frequencies². Such a lag is observed in the normal variations of the seven year light curve of 0954 + 658. Excluding the 1980.95-1981.3 time interval from the seven

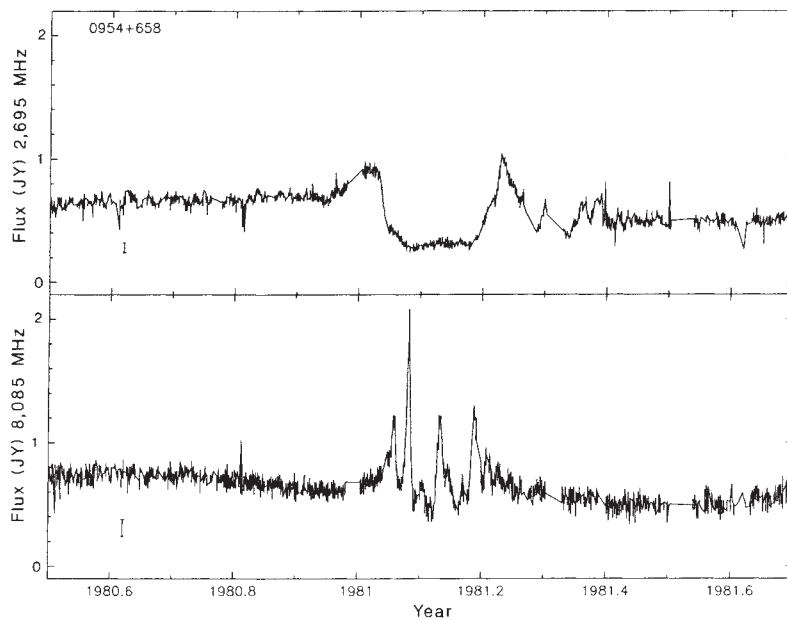


Fig. 1 The 2.7- and 8.1-GHz radio light curves for a 1.2-yr interval surrounding the unusual minimum in 0954+658. Although the individual data points are connected by line segments, none of the peaks shown here, or in Figs 2 and 3, contain fewer than 20 points. The error bars are 2σ in height and were obtained from ref.1.

year radio data we find¹ that the 2.7 GHz variations lag those at 8.1 GHz by approximately 50 days, consistent with an evolving emitter intrinsic to the source. Clearly, the unusual variations in 0954+658 cannot be related to the intrinsic effects that result in a lag between the two observing frequencies.

Rather, the unusual variations appear to be indicative of an occultation event. The rapid decline of the 2.7 GHz flux density just before the period of low, constant flux, and the subsequent rapid rise provides information on the proper motion of the occulter relative to the line of sight. From the FWHM (full width at half maximum) angular size of 0.39 mas at 5 GHz, given by Pearson and Readhead³, and the 15 day timescale for the fall and/or rise in the light curve, we arrive at a proper motion of ≥ 0.03 mas per day. The inequality is included to allow for a possibly larger angular size at 2.7 GHz. If the measured angular size at 5 GHz is set by normal interstellar diffractive scattering ($\theta \propto \lambda^2$), then a proper motion of ≈ 0.09 mas per day is inferred. If the occulter was at a cosmological distance (including in close proximity to the source itself $\approx 10^3$ Mpc), then the relative apparent transverse velocity between source and occulter would necessarily be $\approx 500 c$ (where c is the speed of light). As this represents a quite large apparent velocity, at least an order of magnitude in excess of that typically inferred for very-long-baseline interferometry (VLBI) components, we consider the possibility unlikely that the occulter is local to the source, or at cosmological distance. If the transverse velocity v between line of sight and occulter is less than c , then the occulter must be closer than about 2 Mpc.

We now consider the possibility that the occulter is local to

our galaxy with a transverse velocity $c < 200 \text{ km s}^{-1}$. This velocity is characteristic of material in the galactic halo. When combined with the estimated proper motion (0.09 mas per day) we obtain a distance limit of $D < 1.3 \text{ kpc}$. Conceivably, the occulter could be much closer to the Earth, having a small relative velocity. This raises the possibility that the transverse velocity may be dominated by the Earth's orbital motion. The immersion and emersion peaks at 2.7 GHz would then be expected to be of different duration, as observed.

From the duration of the minimum we estimate an upper limit to the scale size s of the 2.7 GHz footprint, at the Earth. The product of the velocity $v < 200 \text{ km s}^{-1}$ and the duration of the prolonged minimum, ≈ 60 days, gives $s < 7 \text{ AU}$ ($1.05 \times 10^{14} \text{ cm}$). This dimension should be comparable to the physical size of the occulter.

The irregular fluctuations observed at 8.1 GHz and the more regular variations at 2.7 GHz may be explained using refractive effects through irregularities in ionized gas density in the occulter. From the observed fluctuations at 8.1 GHz, the irregularities must focus strongly, occasionally resulting in multiple imaging and, hence, strong spikes in the light curve. This would require that the typical refractive displacement in the observer's plane, due to each irregularity, be at least half the scale over which the refraction angle decorrelates. The random images would rearrange on the correlation scale for the refraction angle.

The correlation scale may be estimated from the number of spikes at 8.1 GHz and the total angular extent of the occulter. The time between the outermost spikes and a proper motion of

Table 1 Sources with unusual minima

Source	ID	z	$l^{\text{II}} (^{\circ})$	$b^{\text{II}} (^{\circ})$	$\lambda (^{\circ})$	$\beta (^{\circ})$	Time interval	Duration (days)
0954+658	Q	0.368	146	43	123	49	1980.95–1981.30	130
1611+343	Q	1.404	55	46	231	54	1985.32–1985.55	85
1502+106	Q	1.833	11	55	220	27	1979.10–1979.60	180
1502+106	Q	1.833	11	55	220	27	1985.40–1985.90	180
0133+476	BL	0.860	131	−14	42	35	1984.95–1985.10	55
1821+107	Q	1.036	40	11	277	34	1984.10–1984.35	90
1641+399	Q	0.595	63	41	238	61	1982.60–1982.70	35

Unusual minima have been found in these sources¹ and are listed in decreasing order of significance at 2,695 MHz. The top three sources have an immersion peak followed by a prolonged minimum and an emersion peak. (See Figs 1, 2 and 3). The remainder of the list includes sources with minima that do not fit any simple category. The optical IDs are Q for quasar and BL for BL Lac object. The redshifts $z^{9,10}$ and galactic and ecliptic coordinates of these sources are listed along with the time intervals and duration of each event.

Fig. 2 The 2.7- and 8.1-GHz radio light curves for time intervals surrounding the unusual minima in 1502+106 (see the caption to Fig. 1).

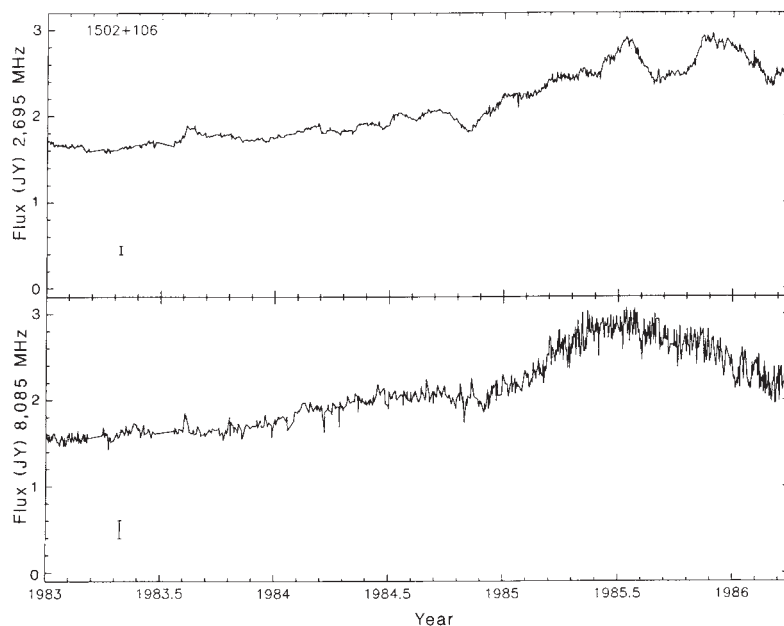
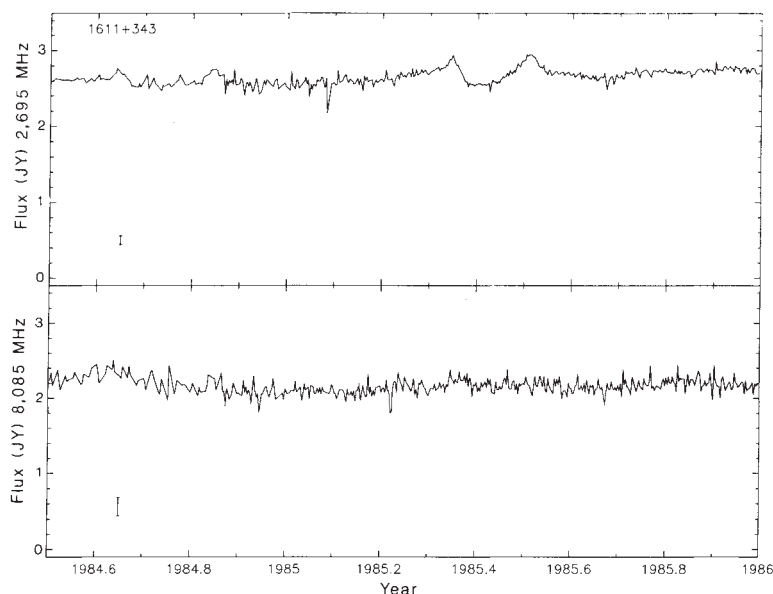


Fig. 3 The 2.7- and 8.1-GHz radio light curves for time intervals surrounding the unusual minima in 1611+343 (see the caption to Fig. 1).



0.09 marc s per day imply an occulter size of ≈ 4.5 marc s and, therefore, a correlation scale of ≈ 1 marc s. The typical refraction angle is then roughly half this amount, or 0.5 marc s. Of course the observer must be near the focal plane to see strong spikes; an unlikely occurrence in general. This may account for the absence of detectable fluctuations at 8.1 GHz in 1502+106 and 1611+343 (see Figs 2 and 3). Table 1 lists all sources from the Green Bank observations¹ that show unusual minima in their radio light curves.

At 2.7 GHz the refraction angle becomes much larger, ≈ 5 marc s, or roughly equal to the angular extent of the occulter itself. A random population of irregularities within the occulting structure will effectively scatter radiation away from the line of sight through the centre of the occulter resulting in a local minimum in the radio light curve. Moreover, when the source is observed outside the occulting object, the normal intensity will be augmented by the scattered radiation producing maxima such as those observed in the 2.7 GHz light curves illustrated in Figs 1-3.

Physically, this model implies inhomogeneities of scale ≈ 0.5 AU, at a distance of $D = 1$ kpc. Since this scale is about 70 times larger than the Fresnel scale, $\sqrt{\lambda D}$, our description of the process as a refractive one is reasonably accurate. The density

fluctuations, n , may be approximated by

$$\theta \approx \frac{2\pi ne^2}{m\omega^2}$$

where e and m are the electron charge and mass respectively, ω is the observing angular frequency and θ is the refraction angle. For $\theta \approx 0.5$ marc s at 8.1 GHz, we obtain $n \approx 4 \times 10^3 \text{ cm}^{-3}$. In this case, the mass of an inhomogeneity is about 10^{19} g. The ionized mass of the occulter would then be $\approx 10^{20}$ g, or possibly more, if it is a linear structure extended along one direction in the sky. Of course, if the irregularities are extended along the line of sight, then the density can be much lower. Also, the preliminary work by Romani *et al.*⁴ on catastrophe theory, as it relates to caustics, may have some application to the physics of the observed events.

As we have seen three events spanning 395 days (see Figs 1-3 and Table 1) in about 160 source-years, a given source is being occulted about $\sigma \approx 7 \times 10^{-3}$ of the time. Although, this is a very crude estimate of the covering factor σ of occulters in the sky, the obvious implication is that there may be a large number of these objects in the galaxy. If the occulters fill the galactic halo (of assumed radius $r = 15$ kpc), then their number density must be $\eta \approx 120 \text{ pc}^{-3} f^{-1}$. The dimensionless factor, f , is included to

allow for the possibility that the occulters may be filamentary, and is the ratio of the linear extent to the width. Note that the inferred number density of occulters is approximately 3 orders of magnitude greater than the number density of stars in the galactic disk 0.12 pc^{-3} , thereby making stellar occultations an unlikely possibility. In the derivation of η we assumed a uniform density distribution $\eta(\text{pc}^{-3})$ of occulters with radius $R(\text{AU})$ filling a volume of radius r (kpc), where $\sigma \approx 8 \times 10^{-8} \eta r R^2$. We have assumed $R \approx 7 \text{ AU}$. Despite the large number of occulters, the associated mass density (averaged over large volumes) is not large, about $10^{-11} M_{\odot} \text{ pc}^{-3}$, using our model for the 0954 + 658 event. The total mass of occulters in the galaxy could be of order $\approx 100 M_{\odot}$. Clearly these structures may involve just a minuscule fraction of the mass of the interstellar medium.

A requirement of the refractive scattering model discussed above, therefore, is the presence of approximately $10^2 \text{ pc}^{-3} \text{ AU}$ -sized objects with ionizing temperatures. There are many important questions concerning the nature and stability of these structures that we are unable to answer at present. Are they in pressure equilibrium with their surroundings? Are they in a state of radiative equilibrium, or do they recombine on short timescales? The nature of these objects still needs to be elucidated by observations of additional events.

We examined the Polaris Observatory Sky Survey prints (POSS), the Infrared Astronomical Survey point source catalogue⁵ (IRAS), and high velocity cloud surveys^{6,7} to look for unusual features along the line of sight to those sources listed in Table 1. From the estimate of the proper motion derived above we expect features in these surveys to be coincident to within one arc s of the source in question. An examination of the POSS shows each source to be situated amongst field stars. There are no extended features in any of these fields.

An examination of the IRAS point source catalogue reveals no infrared point sources within four arc min (eight IRAS beam widths) of the radio positions of the sources listed in Table 1, except for 1641 + 399, which has been identified to have an infrared counterpart. The HVC surveys, like POSS and IRAS, also show no interesting emission features generally coincident with the radio positions of sources with unusual minima. One exception is 0954 + 658, in that it is along the line of sight through the edge of the high velocity cloud A-complex⁶ in the northern galactic hemisphere. The absence of strong evidence for occulters in these surveys is not unexpected given that the inferred angular size of the occulters is of the order of a marc s.

Observations sensitive to microstructure in the interstellar medium have been limited and invariably involve scattering. Hence, it is not completely surprising to find new structure on these scales. Also, identifying unusual signatures in radio light curves has in the past been extremely subjective, rendering interpretation difficult⁸. Now, with the daily monitoring programme at Green Bank, we have sufficient time resolution to unambiguously select features in radio light curves that do not fit the usual behaviour of compact extragalactic radio sources. Evidently, the occulting objects represent a new astrophysical phenomenon, whose nature still needs to be elucidated by observations of additional events. Additional detections of occultation events will initiate numerous radio and optical line and continuum measurements to better determine the physical characteristics of the individual occulter.

Received 22 October 1986; accepted 3 February 1987.

1. Fiedler, R. L. *et al. Astrophys. J. Supp. Ser.* (submitted).
2. Kellermann, K. I. Pauliny-Toth, I. I. K. A. *Rev. Astr. Astrophys.* **19**, 373–410 (1981).
3. Pearson, T. J. & Readhead, C. S. *Astrophys. J.* **248**, 61–81 (1981).
4. Romani, R. W., Blandford, R. D., Goodman, J. J. & Narayan, R. *Bull. Am. astr. Soc. AAS* **18**, 1031 (1987).
5. Beichmann, C. A., Neugebauer, G., Habing, H. J., Clegg, P. E. & Chester, T. S. (eds) *IRAS Catalogs and Atlases* (JPL publication D-1855, Pasadena, 1985).
6. Verschuur, G. L. A. *Rev. Astr. Astrophys.* **13**, 257–293 (1975).
7. Hulsbosch, A. N. M. *Astr. Astrophys.* **340**, 1–25 (1975).
8. Epstein, E. E., Fogarty, W. G., Mottmann, J. & Schneider, E. *Astr. J.* **87**, 449 (1982).
9. Argue, A. N. *et al. Astr. Astrophys.* **130**, 191 (1984).
10. Lawrence, C. R., Pearson, T. J., Readhead, A. C. S. & Unwin, S. C. *Astr. J.* **91**, 494 (1986).

First dynamic spectra of stellar microwave flares

T. S. Bastian* & J. A. Bookbinder†

* Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA

† Joint Institute for Laboratory Astrophysics, Boulder, Colorado 80309, USA

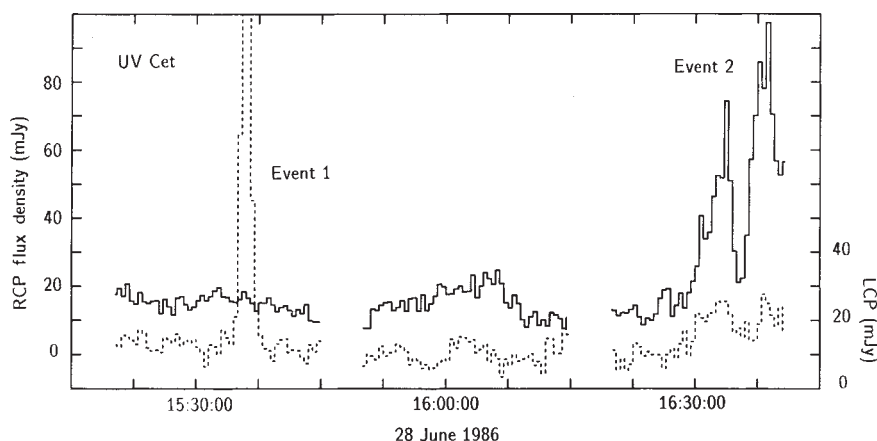
Records of radiation intensity as a function of time and frequency (dynamic spectra) have been profitably used as a probe of the Sun's corona for many years. Motivated by the possibility of using the frequency domain to constrain theoretical models of coherent microwave emission from red dwarf (dMe) flare stars, we have used the Very Large Array in spectral-line mode at 1.4 GHz to obtain the first dynamic spectra of stellar sources other than the Sun. Two very intense, highly circularly polarized, microwave outbursts were observed on the dMe flare star UV Cet (L726-8B), in addition to a slowly varying, unpolarized component. One outburst was purely left circularly polarized and showed no variations as a function of frequency across the 41 MHz band, whereas the other was as much as 70% right-circularly polarized and showed distinct variations with frequency. Although the slowly varying emission is probably due to incoherent gyrosynchrotron emission, the two flaring events are the result of coherent mechanisms. We interpret the coherent emission in terms of plasma radiation and the cyclotron maser instability.

We used the Very Large Array (VLA) on 28 June 1986 in 16 channel spectral line mode centred at 1,415 MHz. The 50-MHz bandwidth was divided into 15 channels of 3.125-MHz width; the sixteenth channel was reserved for recording the integral flux over the inner 37.5 MHz of the bandwidth. The array was divided into two subarrays, one recording right circular polarization (RCP) and the other left circular polarization (LCP). With this configuration we obtained a flux density measurement every 5 s, in each of 15 channels, in both senses of circular polarization. Four dMe flare stars—UV Cet, EQ Peg, AD Leo and YZ CMi—were observed for approximately five hours each. Calibration of the data was performed off-line and a bandpass correction was applied to all antenna baselines. Following standard practice, the channels at either end of the frequency band were excluded from further analysis. Although a sequence of maps was made of the integral flux to confirm the origin of the time varying radiation observed, it was clearly impractical to produce such maps for each 5-s time step, each frequency channel and each polarization. Instead, once the radio emission was confirmed as originating from the source of interest, a direct Fourier transform of the measured complex visibilities was performed as a function of time for each frequency channel and polarization.

Here we describe the slowly varying activity and two episodes of intense flaring activity on UV Cet (L726-8B). During the ~5 hours of observing, we recorded a relatively intense ($S_{\nu} \sim 2\text{--}18 \text{ mJy}$), slowly varying (on a timescale of ~2 hours), broadband, unpolarized continuum. Past reports of the quiescent emission of UV Cet^{1,2} have placed its intensity at ~1 mJy. The companion of UV Cet (L726-8A) was not detected.

Superposed on the slowly varying component (see Fig. 1) were two intense outbursts of quite different character. Event 1 (see Fig. 2) is distinguished by: (1) its simple light curve characterized by a rapid rise (~10 s) followed by a more gradual decay, although two peaks are present near maximum; (2) its short duration—the flux density is above 100 mJy for only 50 s; (3) its large peak flux density of 217 mJy, the most intense microwave burst yet observed from a dMe flare star; (4) its 100% left-hand circular polarization; (5) the broadband nature of its emission, with all frequency channels showing identical light

Fig. 1 The RCP and LCP intensity of the flare star UV Cet at 1.4 GHz over a 37.5 MHz bandwidth is shown for ~1.5 h. The solid curve represents RCP and the dashed curve represents LCP. The LCP curve is offset from the RCP light curve by 10 mJy for clarity. Note that the intensity scale extends only to 100 mJy, thereby cutting off event 1 (see Fig. 2). The differences between event 1 and event 2 with respect to circular polarization, intensity, duration and overall morphology are clearly visible. Gaps in the intensity record are due to calibrator scans.



curves to within the noise (see below). In contrast, event 2 (see Fig. 3), an hour later, is quite different: (1) it shows two high-intensity RCP peaks (labelled A and B in Fig. 3) that are characterized by a relatively slow rise (3 min) followed by a more rapid decay (1.5 min), and complex variability on time-scales as short as the 5-s integration limit; the LCP light curve has low-intensity features that correspond to the RCP peaks; (2) the duration of the entire event is somewhat uncertain because VLA coverage ended when the flux was still high and varying; (3) RCP peaks A and B attain flux densities of 80 and 100 mJy, respectively; (4) both peaks display peak degrees of circular polarization (ρ_c) in excess of 70% in the right-hand sense; each, however, shows large variability in ρ_c with values ranging from 10 to 70% over short times; (5) the emission displays structure across the 41-MHz bandwidth in the form of narrow-band components ($\Delta\nu/\nu \sim 0.002$) present in the emission.

We assessed the reality of the narrowband components in the following way: adopting the premise that the emission is broadband as our null hypothesis, we averaged the emission over all frequency channels and subtracted the result across the frequency band. For broadband emission, the residuals that result from the above procedure display a normal noise distribution. On the other hand, if the emission is 'spiky' in nature, the procedure results in a distribution that is over-represented in the wings; that is, the residuals display significant kurtosis. Two nonparametric tests were applied to the data to determine whether the null hypothesis is valid. The first was a Kolmogorov-Smirnov (KS) two-sample test, where we tested the residuals against a noise distribution resulting from the same subarray of VLA antennas. The second was a kurtosis test. The KS test demonstrated that whereas the residuals of the first event were not distinguishable from a pure noise distribution, the residuals of the second event could not be the result of random noise fluctuations with a confidence level $>90\%$. Second, the first event did not yield residuals with a significant kurtosis; those of the second event did, again with a confidence level $>90\%$.

Several general remarks can be made. First, the few previous microwave outbursts observed from the UV Cet system have been of RCP, leading to the suggestion³ that microwave outbursts from this and other dMe flare stars possess an intrinsic 'handedness'. In view of event 1, this now appears to be unlikely. Second, the brightness temperature of the 1.4 GHz radiation from UV Cet is given by $T_B = 1.3 \times 10^9 S_\nu r_s^{-2}$, where S_ν is the flux density in mJy and r_s is the source radius in units of the stellar radius, $r_* \approx 1.7 \times 10^{10}$ cm. The slowly varying, unpolarized component ranges from $T_B = 2.5 \times 10^9$ K to 2.3×10^{10} K, assuming a constant equivalent source size of radius r_* . If we assume a larger source size, T_B is smaller. The properties of the slowly varying component of the microwave emission from UV Cet are consistent with incoherent gyrosynchrotron emission (that is, emission by mildly relativistic electrons) although the inferred T_B may be near the upper limit for this mechanism⁴. Third, the

maximum T_B (LCP) for event 1 is 2.7×10^{11} K and the maximum T_B (RCP) for event 2 is 1.9×10^{11} K, again assuming $r_s \sim r_*$. In the light of the high degrees of circular polarization observed, these extremely high values of T_B indicate emission by a coherent mechanism.

There are two coherent emission mechanisms believed to be important in the solar corona at radio wavelengths: plasma radiation and the cyclotron maser instability. Both might be expected to be important in producing the coherent microwave emission observed from dMe flare stars. In the case of plasma emission, radiation is produced at either the fundamental or the second harmonic of the electron plasma frequency, $\omega_{pe} = (4\pi n_e e^2 / m_e)^{1/2}$, where n_e is the electron density, e the electron charge and m_e the electron mass. The polarization properties of the radiation are such that fundamental plasma emission can be completely polarized and is in the sense of the O mode, whereas harmonic emission is moderately polarized and can be either X or O mode. In the presence of a non-negligible magnetic field, $\mathbf{B}$, which is probably the case in the corona of a dMe flare star⁵, the frequency is modified as $\omega = (\omega_{pe}^2 + 3k^2 V_e^2 + \Omega_{Be}^2 \sin^2 \theta)^{1/2}$ where k is the wave number, V_e is the electron thermal velocity, and θ is the angle between $\mathbf{k}$ and $\mathbf{B}$. The maximum brightness temperature feasible⁶ for both fundamental and harmonic plasma emission is of the order of $T_B \sim 10^{15}$ K.

For the cyclotron maser, amplification of the magnetoionic X, O or Z modes at the fundamental of the gyrofrequency, $\Omega_{Be} = eB/m_e c$, produces radiation that cannot escape the source because of strong gyroresonant absorption at the second harmonic level. However, direct amplification of the second or higher harmonics of X or O mode radiation, or coalescence of second harmonic Z mode⁷ may result in radiation that can escape the source. This radiation, therefore, can be polarized in either sense. Individual maser pulses can reach extremely high brightness temperatures⁸, $T_B \sim 10^{20}$ K.

Both the cyclotron maser instability and plasma radiation can produce narrow band emission. For a single maser pulse $\Delta\nu/\nu \ll 0.01$. However, because the maser pulses presumably originate over a range of heights and therefore, magnetic field strengths, the overall frequency range produced can be quite large⁴. When plasma radiation is driven by Langmuir turbulence excited by the passage of an electron beam with velocity U_b , the fundamental plasma radiation bandwidth scales as V_i/U_b , where V_i is the ion thermal velocity, leading⁶ to $\Delta\nu/\nu \sim 0.005$. The harmonic plasma radiation bandwidth scales as V_e^2/U_b^2 , where V_e is the electron thermal velocity, yielding $\Delta\nu/\nu \approx 0.01$.

If event 1 is the result of fundamental plasma radiation, a source of Langmuir turbulence is required to drive the emission. Electron beams and shock waves, ubiquitous in astrophysical plasmas, produce 'bump-on-the-tail' or similar distributions which are unstable to the production of Langmuir waves. Such is the case in the solar corona where electron beams and shocks propagating through the corona are known to produce

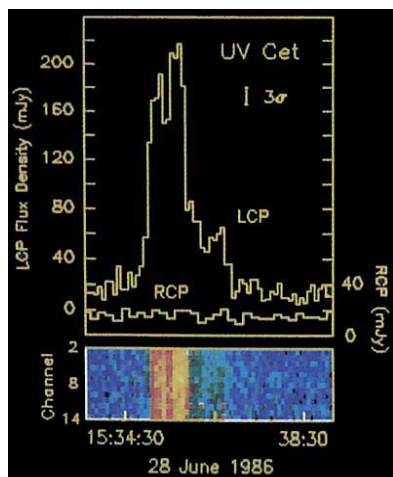


Fig. 2 The upper panel shows the RCP and LCP flux densities ($1 \text{ mJy} = 10^{-26} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1}$) of UV Cet as a function of time (UT) for the inner 37.5 MHz of the frequency band. The LCP flux was averaged every 5 s and the RCP flux every 10 s. The RCP flux scale (right-hand axis) has been offset by -20 mJy . An error bar three times the r.m.s. error ($1\sigma = 4.5 \text{ mJy}$) of the LCP flux is shown. The dynamic spectrum of the LCP radiation of event 1 is displayed in the lower panel. Each pixel represents 5 s of time and one channel of 3.125 MHz bandwidth. The frequency ranges from 1,396.250 MHz (channel 2) to 1,436.875 (channel 14). The r.m.s. error of each pixel is typically 17 mJy. Because intensity differences are extreme in various parts of the band, colour has been used to represent the data.

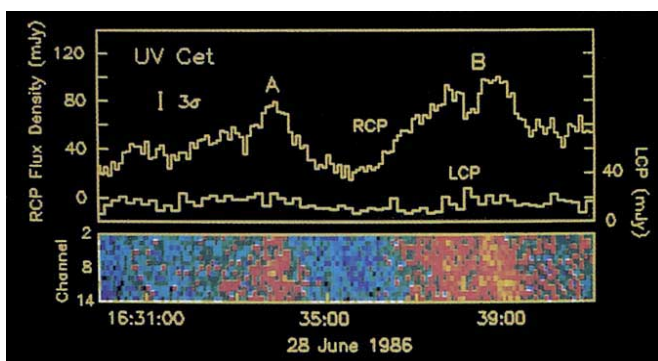


Fig. 3 Same as Fig. 1, but for event 2. Here the dominant sense of circular polarization has reversed. Therefore the RCP flux was averaged every 5 s and the LCP flux every 10 s; the r.m.s. error of the RCP flux is 4.1 mJy. The vertical scales are identical to those of Fig. 1. The dynamic spectrum is of the RCP radiation and has a typical r.m.s. error of 16 mJy.

fundamental and harmonic plasma radiation. The frequency of the emission produced in this way drifts as the disturbance traverses the medium because the electron number density decreases with radius. The drift rate is $\dot{\nu} \sim 0.5 U_b \nu / H_g$, where H_g is the density scale height. Using $H_g \sim 10^{10} \text{ cm}$ and $U_b \sim 0.2c$ (a typical value for an electron beam in the solar corona) a frequency drift rate $\approx 400 \text{ MHz s}^{-1}$ is obtained, which is far too rapid to be resolved by the 5 s integration time used for the observations reported here; hence the emission appears to be broadband. Similarly, for $U_b \sim 1,000 \text{ km s}^{-1}$ (a typical value for a solar shock wave), $\dot{\nu} \sim 6 \text{ MHz s}^{-1}$, still too rapid to be resolved. There is a question regarding whether fundamental plasma radiation can escape the source because of free-free absorption by the overlying plasma. The absorption coefficient is given approximately by $\kappa_\nu \approx 0.2 n_e^2 T^{-3/2} \nu^{-2}$, where T is the temperature of the ambient plasma and $\nu = \omega/2\pi$ is the frequency of the radiation. The optical depth due to thermal absorption is therefore $\tau \approx 0.1 n_e^2 \nu^{-2} T^{-3/2} H_g$, assuming $n_e = n_o \exp(-h/H_g)$. Assuming the plasma radiation is near the maximum feasible brightness temperature ($T_B \sim 10^{15} \text{ K}$) even a value of τ as large as 10 allows sufficient radiation to escape the source. With $T = 10^7 \text{ K}$ and $n_o = \pi m_e \nu^2 / e^2 = 2.5 \times 10^{10} \text{ cm}^{-3}$, the value $\tau = 10$ occurs with a density scale height of $\sim 10^{10} \text{ cm}$, which is entirely plausible. If the radiation detected from event 1 was plasma radiation at the fundamental with $T_B \sim 10^{15} \text{ K}$, a relatively compact source size of radius $3 \times 10^8 \text{ cm}$ is implied. It is conceivable that event 1 is due to the cyclotron maser instability. Here, too, extremely high degrees of circular polarization are possible. Although individual maser pulses are narrow band and of short duration, if they occur with uniform density over the frequency band with a duty cycle $\ll 5 \text{ s}$, the 5-s time constant used for the observation will average them to produce a smooth pseudo-continuum.

Event 2 is different from event 1 in all its observed properties. The light curve is qualitatively similar to spike burst mor-

phologies observed on the Sun⁹ and other dMe flare stars^{10,11}. In particular, it displays intense, 'spiky' emission in one polarization and smoothly varying, low-intensity emission in the other polarization (best illustrated by Fig. 1). In these respects, the burst appears to be cyclotron maser emission operating in RCP against a background of incoherent gyrosynchrotron emission. If such is the case, a magnetic field of $B \sim 250$ gauss is implied, assuming that the radiation is at the second harmonic. Unfortunately, in lieu of higher time resolution and/or a larger relative bandwidth, we cannot positively identify the emission mechanism. It is clear, however, that if event 1 and event 2 are due to the same coherent emission mechanism, the physical circumstances from which each arose were quite different.

The National Radio Astronomy Observatory is operated by Associated Universities, Inc. under contract with the NSF. We thank P. Crane, K. Sowinski and A. Pedlar for their invaluable assistance in preparing the observational program on which this work is based, and M. McKean for help with calibration. In addition we thank G. Dulk and J. Linsky for a critical reading of the manuscript, and D. Gary and A. Rots for useful discussion and comments. This work is supported by NASA under grants NSG-7287, NAGW-91, and NGL 06-003-057 to the University of Colorado.

Received 23 October 1986; accepted 27 January 1987.

1. Kundu, M. & Shevgaonkar, R. K. *Astrophys. J.* **297**, 644-648 (1985).
2. Fisher, P. L. & Gibson, D. M. *Astrophys. J.* (in the press).
3. Gibson, D. M. in *Solar Terrestrial Physics Proc. Sec. Indo-US Workshop* (eds Kundu, M., Biswas, B., Reddy, B. M. & Ramadurai, S.) 43-57 (National Physical Laboratory, New Delhi, 1986).
4. Dulk, G. A. *A. Rev. Astr. Astrophys.* **23**, 169-224 (1985).
5. Saar, S. & Linsky, J. L. *Astrophys. J.* **299**, L47-50 (1986).
6. Melrose, D. B. *Plasma Astrophysics Vol. 2* (Gordon & Breach, London, 1980).
7. Winglee, R. M. *J. geophys. Res.* **90**, 9663-9674 (1985).
8. Melrose, D. B. & Dulk, G. A. *Astrophys. J.* **259**, 844-858 (1982).
9. Slottje, C. *Nature* **275**, 520-521 (1978).
10. Lang, K. R., Bookbinder, J. A., Golub, L. & Davis, M. M. *Astrophys. J.* **272**, L15-18 (1983).
11. Holman, G. D., Bookbinder, J. & Golub, L. in *Radio Stars* (eds Hjellming, R. M. & Gibson, D. M.) 35-38 (Reidel, Dordrecht, 1985).

High-resolution imaging of the M87 core

William G. Bagnuolo Jr* & Richard W. Chambers†

* Center for High Angular Resolution Astronomy,
Georgia State University, University Plaza, Atlanta,
Georgia 30303, USA

† The Aerospace Corporation, PO Box 92957, Los Angeles,
California 90009, USA

The nature of the centre of the elliptical galaxy M87 is still a matter of controversy. To obtain improved spatial resolution we have recorded alternately a number of short-exposure images of the M87 core and several comparison stars with an ISIT camera at the AMOS 1.6-m telescope at Haleakala, Maui. An analysis of selected images (0.54–0.60 arc s FWHM (full width at half maximum)) indicates that there is no significant difference between the apparent widths of the M87 core and the stars. The true core width appears to be definitely <0.3 arc s and probably <0.20 arc s. Thus, we do not observe a hypothetical stellar cusp created by the potential of a massive black hole.

Among other interesting features, such as the 'jet', M87 is known to have a bright luminosity 'spike' at its centre, unresolved at 1 arc s resolution. In 1978, Sargent *et al.*¹ presented spectroscopic data that showed a rise in the velocity dispersion (ΔV) from $r = 10$ to $r = 1.5$ arc s from centre. Combined with the photometric imaging data of Young *et al.*², the results suggested that a central black hole of $5 \times 10^9 M_\odot$ existed.

More recently, however, Dressler's³ spectroscopic data at average distances $r = 13$ and $r \sim 0.35$ arc s with higher spatial resolution showed that ΔV does not rise as rapidly at the centre as was thought. The results also showed that although a non-thermal continuum is strong in the blue region, most of the light in the visible and red regions comes from starlight. Several models were now compatible with these data, including a compact self-gravitating star cluster (diameter ~ 0.1 arc s) or a larger distribution of stars caught in a black-hole cusp (diameter ~ 0.4 arc s)⁴⁻⁵. The images of the core by Young *et al.*, ~ 1.1 arc s FWHM, were unable to resolve the issue despite enhancement⁶.

The project was begun in 1982 to attempt to provide better resolution for the M87 core (the luminosity spike) with the 1.6-m telescope at Haleakala, a site with excellent seeing when the inversion layer is low. Short exposures (1/4–1/8 s) were used to gain an additional advantage in resolution⁷.

Although some data under fair conditions were obtained in 1983, useful data under good seeing conditions (about 0.8 arc s FWHM) were obtained on 12 April 1985. Galaxy and comparison stars were observed alternately for about 35–40 s per run. The latter, whose magnitudes of 14.7–17.2 bracketed the galaxy, were listed by T. Boroson (personal communication, 1984) who also generously provided their B, V, and R magnitudes. The passband was an S–20 ER, centred at about 5,600 Å (thereby avoiding most of the blue non-thermal continuum).

The star and galaxy core observations were done in about the same location on the ISIT camera field of view (to ± 15 pixels)

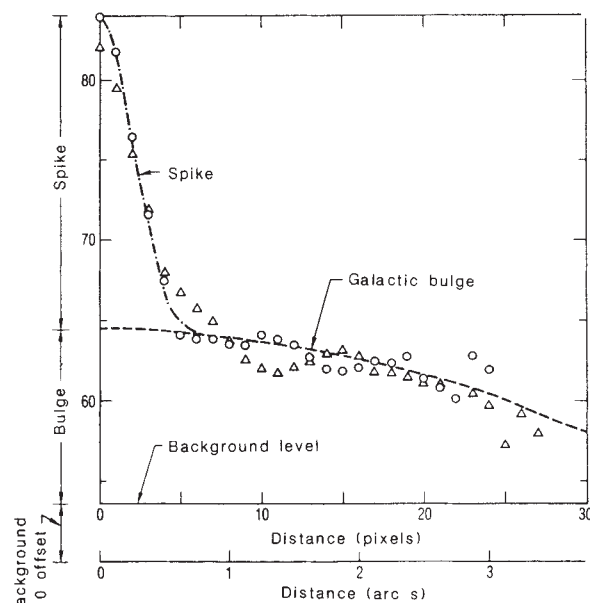


Fig. 1 Galaxy core (luminosity spike) and bulge, showing resolution of the former.

near pixel (x, y) location (190, 250), that is, offset left of centre to avoid the locally high incidence of ion events at the centre. During these observations the galaxy was near the meridian, within 12° of the zenith. A Barlow lens was added to magnify the telescope focal ratio of $f/15$ by $\sim$ two.

The output of the ISIT camera was recorded onto a PAL 1-in width videotape. Because 1-in PAL video machines were not available, and because of digitizer incompatibility, the tape was remastered into 3/4-inch NTSC format at Image Transforms, Inc. (North Hollywood, California). (Thus, pixel size increased by ~ 6.5 .) The data, played by a 3/4-inch VCR with a time-base corrector, were digitized with a PC-Vision 8-bit digitizer with an IBM-PC XT host computer. Consecutive 32×32 pixel centred frames were stored in a 2 mega byte memory. The x, y plate scales were 0.147, 0.133 arc s per pixel based on CCD frames of globular cluster M3 by H. Cohn.

Table 1 shows the colours of the comparison stars, their estimated in-band intensities and the observed intensity. Three out-of-focus images of SAO82950 were used to check the linearity at the lowest light levels. The relative actual intensities were estimated from the size of the images. The results show that the detector is approximately linear (to $\pm 5\%$) over the intensity range of interest.

Flat-fielding was done at three intensity settings by facing the telescope at the dome cover and turning on the floor lights. The 64×32 pixel region where the galaxy and star observations took place was uniform in response to 2.0% r.m.s. and, therefore further corrections were not made.

FORTTRAN routines were devised to compare the composite intensity profiles of the core and the comparison stars from the centred 'best' 5% and 10% of the images. The criterion for selecting the 'best' images is the r.m.s. spread of the pixels in a background-subtracted picture that are brighter than half the average of the two brightest pixels. We have experimented with other criteria, such as the half-widths in the y -direction, with similar results.

The main data reduction program also includes a routine to subtract the M87 galactic 'bulge', which has about half the core's in-band intensity. Figure 1 shows the intensity for two individual galaxy images, the best of a sample of 64 images. It was found that assuming the bulge is just a flat background led to an increase of 0.03 arc s in the measured FWHM of the core.

Table 1 Comparison star data

Star no.	Position offset from M87 core (arc s)		V	B–V	V–R	S–V (calculated)
	RA	Dec				
6	109.5	–117.9	14.69	1.01	0.60	0.22
7	114.9	+62.0	15.54	1.37	0.93	0.20
8	–165.0	+52.8	15.80	0.75	0.60	0.19
9	179.6	+92.6	17.16	0.86	0.63	0.20
M87 core			16.3	0.73	0.59	0.19
M87 (total)			15.9	0.87	0.74	0.20

Table 2 Star and galaxy Profiles: (1/4 s, best 10% of images)

	Pixel	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
Star	7	0.050	0.093	0.150	0.227	0.370	0.603	0.867	1.00	0.883	0.600	0.343	0.207	0.130	0.083	0.047
	8	0.060	0.097	0.152	0.230	0.352	0.575	0.853	1.00	0.862	0.548	0.307	0.183	0.113	0.077	0.050
	9	0.043	0.077	0.107	0.173	0.310	0.577	0.873	1.00	0.820	0.553	0.317	0.213	0.147	0.117	0.073
Galaxy	7	0.050	0.115	0.185	0.245	0.335	0.550	0.840	1.00	0.855	0.540	0.295	0.210	0.165	0.095	0.045
	8	0.048	0.068	0.121	0.192	0.325	0.562	0.865	1.00	0.845	0.547	0.315	0.186	0.115	0.095	0.078
	9	0.057	0.109	0.184	0.270	0.390	0.603	0.873	1.00	0.877	0.600	0.350	0.189	0.104	0.070	0.060
Star	(all)	0.053	0.091	0.140	0.215	0.346	0.582	0.862	1.00	0.857	0.562	0.318	0.197	0.126	0.088	0.055
Galaxy	(all)	0.051	0.090	0.153	0.225	0.344	0.569	0.861	1.00	0.856	0.558	0.319	0.193	0.125	0.089	0.065

Table 3 Galaxy core versus star FWHM (top 10% of images)

	Star no.	Galaxy core width	N_g	s.d. _g	Star width	N_s	s.d. _s
	7	0.567	2	0.084	0.630	3	0.010
	8	0.584	5.5	0.032	0.592	6	0.028
	9	0.632	4	0.077	0.571	3	0.010
Average		0.604	11.5	0.055	0.603	12	0.028
Difference		0.001 ± 0.018					
Difference	(star 8 only)	-0.008 ± 0.018					
Composite profile	(table 2)	0.596			0.597		

Table 4 Galaxy-core significance

Intensity cutoff	Core width	Star width	s.d.	Core-star (sign.)	Core-0.25 (sign.)	Core-0.30 (sign.)	Core-0.40 (sign.)
0.5 (FWHM)	0.596	0.597	0.018	-0.001 (-0.05)	-0.082 (-4.55)	-0.111 (-6.2)	-0.179 (-9.9)
0.7	0.397	0.396	0.014	0.001 (0.07)	-0.071 (-5.1)	-0.094 (-6.7)	-0.144 (-10.3)

Before stating some of the preliminary results, we note the following. First, we have concentrated on comparisons of the galaxy with stars 8 and 9, particularly the former as it is closest to the M87 core in intensity (see Table 1). Secondly, the azimuthal oil bearings of the telescope allowed a wind-induced wobble of up to 1 arc s peak-to-peak predominantly in the x-direction (inclination 10–15°) with a period of about 1.5 s. For this reason and others, such as better pixel resolution in the y-direction, the half-width measurements were made in the y-direction. Thirdly, the star or galaxy-core profiles in the y-direction were made 5 pixels wide across the centroid, rather than 1 pixel wide, to improve the signal-to-noise ratio. This results in a slight over-statement of the FWHM.

Finally, we have concentrated on the results for the best 10% of the 1/4-s exposures, because a somewhat better signal-to-noise ratio could be obtained in that way. (The increase in the number of frames tends to outweigh the gain in resolution.) The results of the 1/8-s exposures and/or best 5% of the frames were essentially the same and will be presented later.

Table 2 shows the average of the star profiles and the corresponding galaxy profiles for the best 10% of the 1/4-s exposures (for example, 'galaxy 8' refers to the galaxy-core profiles taken in comparison with star 8). Note the basic similarity of the star profiles, which is another check of system linearity. The bottom of Table 2 compares the composite star and galaxy profiles (stars 7, 8, 9 given weights 1:2:1), which are quite similar.

Figure 2 plots the composite star and galaxy profiles. The composite star profile was fitted by a gaussian and lorentzian profile fit via a nonlinear least-squares procedure, based on ref. 8. The fitted curve was convolved with a gaussian profile with a FWHM of 0.4 and 0.25 arc s. These curves are both considerably larger than the galaxy core profile.

The profiles from individual runs were analysed to obtain an estimate of the uncertainties in the composite profile widths. Table 3 presents the galaxy-core width for each star comparison, the number of observations (N) and the estimated standard deviation (s.d.); and the same quantities for the star observa-

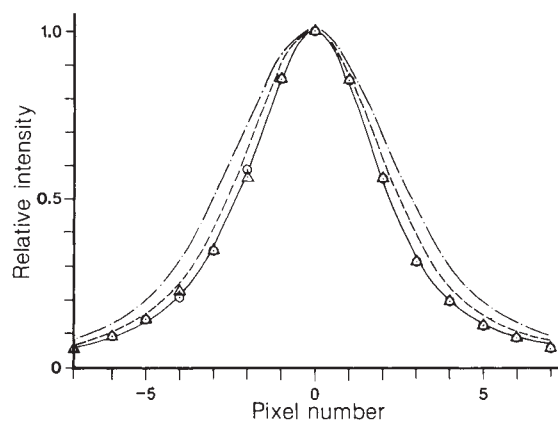


Fig. 2 A comparison of the composite galaxy profile (triangles) and star profile (circles). Dashed line, 0.25-arc s objects; dashed and dotted line, 0.4-arc s objects (0.133 arc s per pixel).

tions. (Note that one galaxy observation went on long enough for another half data set to be collected, also that the first star observation of star 9 was missed.)

Table 4 compares the profiles of the galaxy core to the star and the star convolved with hypothetical gaussian objects with widths of 0.25–0.40 arc s. The values of the differences in width in terms of the estimated s.d. of the core-star width are given in parentheses. Thus, even the difference between the core and the 0.25-arc s object appears statistically significant. It therefore seems unlikely that a core with a true width of 0.25 arc s or larger could have failed to be measurably larger than the star images.

Finally, we briefly consider the effect of varying integration time and frame selection. The FWHMs of the core and star 8, respectively, for the best 5% of the whole data set are 0.540 ± 0.018 and 0.564 ± 0.011 for 1/8-s exposures and 0.554 ± 0.014

and 0.570 ± 0.013 for 1/4 s; for the best 10% they are 0.579 ± 0.014 and 0.567 ± 0.008 for 1/8 s and 0.596 ± 0.010 and 0.587 ± 0.009 for 1/4 s. There are no significant differences between the star and core widths, but the core widths minus the 0.25-arc s object are significant to -4.8 to -6.0σ . There is an additional reduction of about 7% in FWHM for the best-5%, 1/8-s images compared with the best-10%, 1/4-s images. The former are also about 30% smaller than the uncompensated seeing disk (0.8 arcs), a result consistent with those recently obtained at Mauna Kea⁹.

To sum up, the principal conclusion is that the galaxy core is unresolved and less than 0.25 arc s in diameter. These measurements therefore agree with those of ref. 10. Although the large-black-hole model is eliminated, the resolution of the Hubble Space Telescope may be required to verify the existence of a smaller-black-hole cusp. Also, the recent evidence¹¹ of strong emission lines in the spectrum may indicate a larger non-stellar light component than previously thought.

Received 10 September; accepted 29 December 1986.

1. Sargent, W. L. W., Young, P. J., Boksenberg, A., Shortridge, K. & Lynds, C. R. *Astrophys. J.* **221**, 731-744 (1978).
2. Young, P. J., Westphal, J. A., Kristian, J. & Wilson, C. A. *Astrophys. J.* **221**, 721-730 (1978).
3. Dressler, A. *Astrophys. J.* **240**, L11-L16 (1980).
4. Young, P. J. *Astrophys. J.* **217**, 287-295 (1977).
5. Bahcall, J. N. & Wolf, R. A. *Astrophys. J.* **209**, 214-232 (1976).
6. Binney, J. & Mamon, G. A. *Mon. Not. R. astr. Soc.* **200**, 361-375 (1982).
7. Roddier, C. & Roddier, F. *J. opt. Soc. Am.* **65**, 664-667 (1975).
8. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*, 96 (McGraw-Hill, New York, 1969).
9. Christian, C. & Racine, R. *Publs astr. Soc. Pacif.* **97**, 1215-1226 (1985).
10. Lauer, T. & Kormendy, J. *Astrophys. J.* (in the press).
11. Filippenko, A. V. & Sargent, W. L. W. *Astrophys. J. Suppl. Ser.* **57**, 503-522 (1985).

MERLIN polarization observations of the quasar 3C380

C. Flatters*

University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

A model, in which an extragalactic radio source consists of an energy source at the centre of a galaxy, feeding distant radio lobes via two oppositely directed jets of plasma provides a qualitative explanation of the appearance of most sources, any differences between them resulting from orientation dependent effects¹, but does not however, account for the steep-spectrum compact sources (SSCSs) which comprise ~30% of all steep-spectrum extragalactic radio sources^{2,3}. One possibility is that a dense interstellar gas prevents jets leaving the galaxy in these objects⁴. In this case the gas should be detectable through its effects on polarized radio emission. High resolution observations at comparatively low radio frequencies are required to test for this. Here, we present a MERLIN map of 3C380, an SSCS with a structure that may be interpreted as the result of interactions between a jet and gas⁵. The quasar 3C380 is at redshift $z = 0.691$ (ref. 6).

The polarized radio emission from 3C380 was mapped at 1,666 MHz using five telescopes of the Multi-Element Radio-Linked Interferometer Network (MERLIN)⁷ on 28 April, 1984. The map (Fig. 1) was made using a restoring beam of 0.35 arc s, chosen to allow direct comparison with Very Large Array (VLA) maps made at higher radio frequencies. The contours show the total radio brightness while the polarization is represented as superposed vectors. The map is limited by thermal noise with an r.m.s. amplitude of about 1 mJy per beam in all Stokes parameters. There is no sign of any contamination of the linear

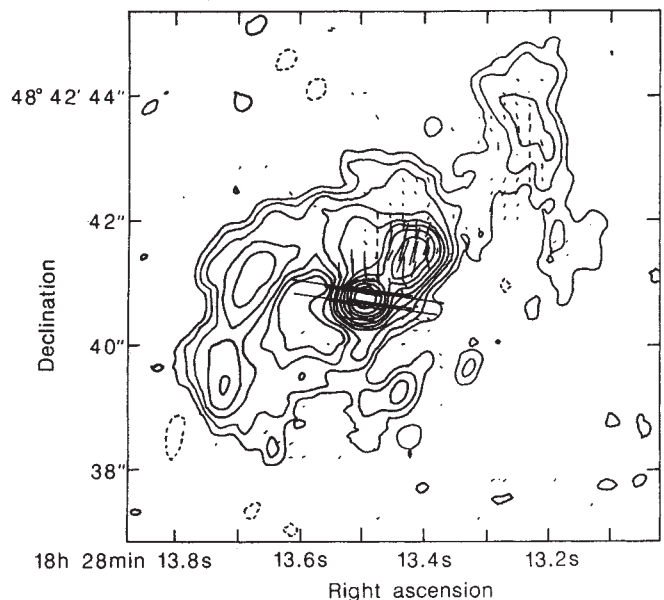


Fig. 1 The MERLIN 18-cm polarization map of 3C380. Contour levels are -4, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2,048 mJy per beam of total intensity. Vectors have directions parallel to the observed electric vector and lengths proportional to polarized brightness. A vector 1 arc s long represents a polarized brightness of 100 mJy per beam. Positions marked on the axes assume that the core lies at the catalogue position of 3C380. Peak flux = 3.76 Jy per beam.

polarization arising from inaccurate subtraction of the instrumental polarization.

The map shows a bright core surrounded by complex structure which has already been described by Wilkinson *et al.*⁵. The core, the jet in position angle -52° (which includes 2 bright knots not separated by the contours used in Fig. 1), and some diffuse structure adjacent to the jet are clearly polarized (Table 1). Comparison with VLA maps at 5 GHz and 15 GHz (P. N. Wilkinson and T. J. Cornwell, unpublished) shows that these are not significantly depolarized at 1,666 MHz (Table 2), although the arc of emission east of the core has depolarized completely at this frequency. The rotation measure is 30 rad m^{-2} everywhere that polarization is detected at 1,666 MHz, with a typical error of $\leq 5 \text{ rad m}^{-2}$. The nearby source 3C381 has a similar rotation measure⁸ and so most of this rotation is probably caused by matter in our own galaxy.

On the basis of total intensity maps alone, Wilkinson *et al.*⁵ concluded that the most likely cause for the complex appearance of 3C380 was that the jet was being distorted by dense gas impinging on it from the south-west. Indeed the appearance of

Table 1 Peak total brightnesses of features in Fig. 1, together with their corresponding polarized brightnesses, fractional polarizations and polarized position angles

Component	Peak brightness (mJy beam ⁻¹)	Polarized brightness (mJy beam ⁻¹)	Fractional polarization (%)	Position angle (deg.)
Core	$3,274 \pm 19$	307	8.3	81.5
Knot A	830 ± 20	23	2.7	-2.9
Knot B	642 ± 60	90	14.0	-14.7
Arc	82	<3	<4	—

The 2 knots in the jet are here designated A (closest to the core) and B. The peak total brightnesses of the core and the knots were determined by fitting Gaussian models to the map, all other quantities were measured by taking point values. Errors from the model fitting are shown for the core and knots.

* Present address: MPI für Radioastronomie, Auf dem Hügel 69, D-5300 Bonn 1, FRG.

Table 2 Fractional polarizations of the components listed in Table 1 as a function of wavelength

Component	2 cm	6 cm	18 cm
Core	1.8%	6.9%	8.3%
Knot A	3.5%	2.8%	2.7%
Knot B	13.5%	11.2%	14.0%
Arc	13–25%	15–20%	<4%

The 2 cm and 6 cm values are derived from VLA maps made by P. N. Wilkinson and T. J. Cornwell.

the source, the curved arc in particular, suggests that the jet is being bent by a rotating disc of gas, as inferred for some Seyfert galaxies⁹. Assuming that the energy radiated by the arc of emission (of order 10^{37} W for H_0 (Hubble constant) = $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 0.5$) is supplied by the jet and that jet kinetic energy is transformed into synchrotron radiation with an efficiency not exceeding 10%, the minimum density of external gas required to produce such a bend may be estimated from conservation of momentum arguments. Note that because the jet must carry a fixed amount of energy, the momentum flux of the jet is minimized if the material in the jet flows at the speed of light. If it is assumed that the external gas is not moving at more than $1,000 \text{ km s}^{-1}$ (see velocities of 500 km s^{-1} measured in the SSCS 3C48¹⁰) and that the jet radius is smaller than 1 kpc then a lower limit on the density of external matter of 10^6 electrons per m^3 is obtained. As it is unlikely that a fast jet can be bent suddenly through a large angle without it being disrupted and the external density derived above is inversely proportional to the jet speed this limit should be regarded as highly conservative. Such a large density of thermal matter would have a significant effect on the polarized flux from 3C380 which may be compared with what is observed.

The radio structure seen in Fig. 1 is embedded in a diffuse halo that may be up to 25 arc s in size (P. N. Wilkinson, personal communication). This presumably accounts for the 7 Jy of the total flux density of 3C380¹¹ that are not detected by MERLIN. The equipartition magnetic field of this halo, of the order of 1 nT, provides the best available estimate for the magnetic field in the matter surrounding the structures seen in Fig. 1. A uniform screen of density 10^6 electrons per m^3 and thickness 1 kpc threaded by a magnetic field of this strength would give rise to a rotation measure of $8,000 \text{ rad m}^{-2}$, very much higher than that observed. Although this rotation measure could be significantly reduced by irregularities in the external medium or in the magnetic field within that medium on a scale smaller than the beam size, depolarization would then be expected¹². As no depolarization is observed close to the jet it must be concluded that, interpreted in the standard manner, the observations of linear polarization reported here are not consistent with the hypothesis that 3C380 is distorted by pressure exerted by a moving external gas.

The author acknowledges receipt of an SERC research studentship during the course of this work and thanks P. N. Wilkinson and R. G. Conway for helpful comments on earlier drafts of this paper.

Received 1 October 1986; accepted 5 February 1987.

1. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. Mod. Phys.* **43**, 255–351 (1984).
2. Kapahi, V. K. *Astr. Astrophys. Suppl. Ser.* **43**, 381–393 (1981).
3. Peacock, J. & Wall, J. V. *Mon. Not. R. astr. Soc.* **198**, 843–860 (1982).
4. van Breugel, W. J. M., Miley, G. K. & Heckman, T. M. *Astr. J.* **89**, 5–22 (1984).
5. Wilkinson, P. N., Booth, R. S., Cornwell, T. J. & Clark, R. R. *Nature* **308**, 619–621 (1984).
6. Lynds, C. R., Stockton, A. N. & Livingston, W. C. *Astrophys. J.* **142**, 1667–1673 (1965).
7. Thomasson, P. Q. *J. R. astr. Soc.* **27**, 413–431 (1986).
8. Simard-Normandin, M., Kronberg, P. P. & Butten, S. *Astrophys. J. Suppl. Ser.* **45**, 97–111 (1981).
9. Wilson, A. S. & Ulvestad, J. S. *Astrophys. J.* **236**, 576–594 (1982).
10. Boroson, T. A. & Oke, J. B. *Nature* **296**, 397–399 (1982).
11. Kühr, H., Witzel, A., Pauliny-Toth, I. I. K. & Nauber, U. *Astr. Astrophys. Suppl.* **45**, 367–430 (1981).
12. Burn, B. J. *Mon. Not. R. astr. Soc.* **133**, 67–83 (1966).

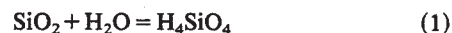
Increased solubility of quartz in water due to complexing by organic compounds

P. Bennett & D. I. Siegel

Department of Geology, Syracuse University, Syracuse, New York 13244, USA

Quartz is the most stable natural solid phase of silica. It weathers extremely slowly at the Earth's surface¹, and often resists weathering even after all other silicate minerals have been degraded. However, there is ample evidence from both ancient and modern environments indicating enhanced dissolution and mobility of silica under conditions that cannot easily be explained by the inorganic controls of quartz solubility². Increased solubility of quartz has been observed particularly in soils rich in organic material; however, no direct link between dissolved organic carbon and dissolved silica has been identified³. Here we present evidence for an increase in the solubility of quartz in a natural water brought about by dissolved organic compounds. These compounds were produced by the biodegradation of petroleum, and consist largely of a complex mixture of organic acids. We propose that silica is being complexed and mobilized by these organic acids in waters having close to neutral pH.

This paper reports on observed evidence of the enhanced aqueous dissolution of quartz in a shallow groundwater system contaminated by crude petroleum. Inorganic controls of quartz solubility are well known. At pH values less than about 9, quartz dissolves in water by the reaction



The aqueous solubility of quartz at 25°C is $3\text{--}5 \text{ mg l}^{-1}$ as Si, whereas in the uncontaminated groundwater at our study site, the equilibrium solubility is $2\text{--}3 \text{ mg l}^{-1}$ ^{4,5}. Many natural non-alkaline waters contain concentrations of dissolved silica well in excess of the equilibrium solubility of quartz⁶. This apparent 'supersaturation' with respect to quartz is often attributed to the relatively rapid hydrolysis of aluminosilicate minerals⁷ compared with the slower precipitation of quartz⁸. However, as long as the activity of dissolved silicic acid exceeds its equilibrium solubility, further dissolution of quartz does not occur. Although the concept of organic complexes of silica has previously been proposed^{9,10}, the solubility of quartz in natural waters has not been directly related to concentrations of dissolved organic carbon³.

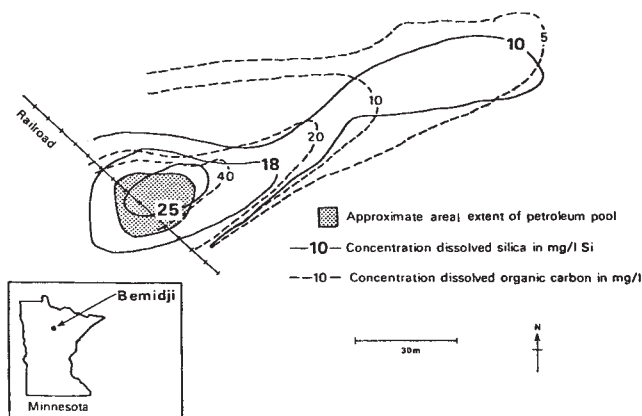
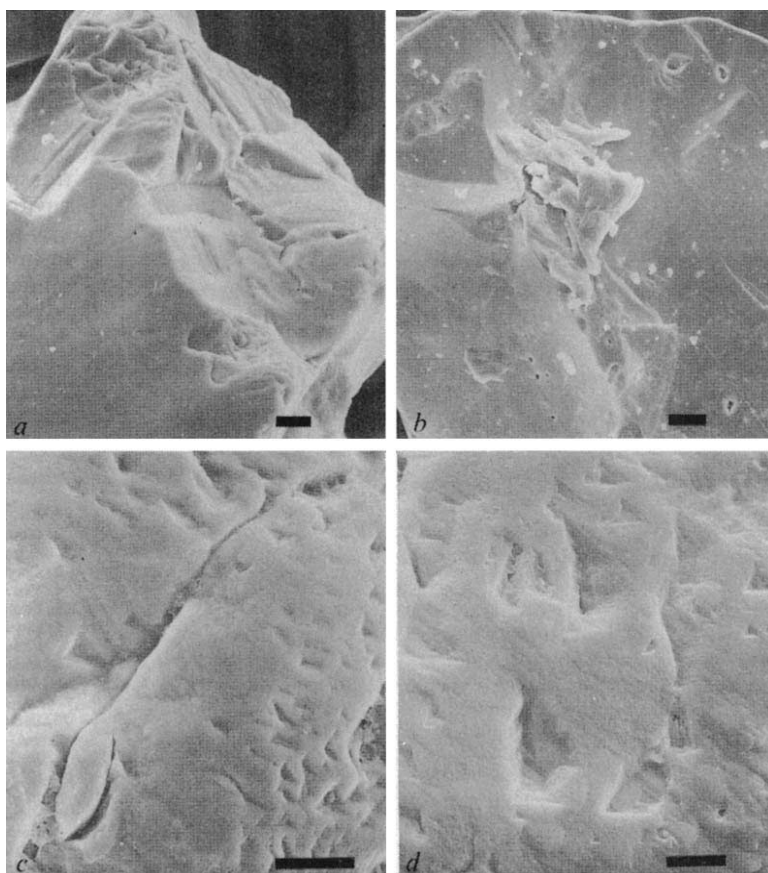


Fig. 1 Site map showing areal extent of petroleum pool, and concentration isopleths of dissolved organic carbon and dissolved silica. Groundwater flow is towards the north-east at a rate of 0.3 to 1.0 m per day .

Fig. 2 *a*, SEM micrograph of quartz sand grain from an uncontaminated well. Note the clean conchoidal fracture surfaces and complete lack of chemical etch pits. These features show a glacial origin with minimal fluvial influence (see text). Scale bar, 10 μm . *b*, SEM micrograph of quartz sand grain collected from within the petroleum pool. Note again the lack of evidence of chemical etching on this grain, suggesting the absence of a direct oil-solid interaction. Scale bar, 10 μm . *c* and *d*, SEM micrographs of quartz sand grains from a contaminated well at the edge of the oil pool in a region of high DOC concentrations. Note the clearly defined triangular etch pits and the generally rough surface of the grains. These features indicate active chemical etching (see text). Scale bar, 10 μm .



We see active chemical etching of quartz under apparent supersaturated conditions in a shallow, glaciofluvially-deposited sand aquifer near Bemidji, Minnesota. The study site was contaminated by crude petroleum from a pipeline rupture in 1979 that resulted in a pool of petroleum floating on the water table¹¹ (Fig. 1). Soil samples from the contaminated and uncontaminated zones were collected using a split-spoon sampler from wells drilled to a level immediately below the water table. Particle-size separation was done by the dry sieve method¹², and the size fraction between 77 and 225 μm was examined using an ETEC Autoscan scanning electron microscope (SEM). Mineral identification was done by energy dispersive analysis of X-rays using a Kevex 4000 spectrometer. Groundwater samples were collected from each well and filtered at 0.1 μm . Concentrations of total dissolved silica were analysed using d.c.

plasma emission spectroscopy. Concentrations of monomeric silicic acid were determined by the silico-molybdate method¹³. Dissolved organic carbon (DOC) was analysed by the wet oxidation method¹⁴. Temperature and pH were measured using a flow-through cell¹⁵. All water samples were also analysed for alkalinity, conductivity and major dissolved solutes (Table 1).

Concentrations of dissolved organic carbon are elevated in the contaminated zone and outline a distinct plume trending down the hydraulic gradient from the oil pool (Fig. 1). Microbial oxidation of the petroleum is active at the site¹⁶ and results in the production of carbon dioxide and various mono- and multiprotic organic acids^{14,16}. These by-products have depressed the pH in the contaminated zone from a background value of 7.7 to as low as 6.7. The concentrations of dissolved monomeric silicic acid and total dissolved silica were found to be nearly identical, and are also elevated in the contaminated zone (Table 1). The concentration gradients of dissolved silica outline a plume similar to the plume of DOC (Fig. 1). Silica was not detected in the original oil by neutron activation analysis.

Quartz grains from background wells surrounding the contaminated zone have surface features diagnostic of grains subjected to mechanical grinding typical of that found in glaciofluvial sediments¹⁷. These grains have clean conchoidal fracture surfaces and cleavage faces, and there is no evidence of chemical weathering on their surfaces (Fig. 2*a*). Similar features are observed on grain surfaces of aluminosilicate minerals.

In contrast, quartz grains from the area of highest DOC have surfaces significantly altered by chemical weathering. In particular, the surfaces are covered with crystallographically oriented triangular etch pits, solution pits and solution channels (Fig. 2*c* and *d*). These features are diagnostic of chemical etching normally associated with very alkaline environments¹⁷. Chemical etching of aluminosilicate minerals, such as feldspars, is also found only in the zone of highest DOC concentrations. The dissolution features are highly unusual and unexpected in

Table 1 Representative water analysis along plume axis

Analysis	Distance down gradient of pipeline rupture (m)				
	(Background)	5	36	90	130
Temperature ($^{\circ}\text{C}$)	12.0	11.0	11.0	11.0	10.0
pH	7.70	6.70	6.90	6.95	7.17
Alkalinity	234	720	677	506	466
Si	9.47	28.8	16.3	9.89	9.94
Al ³⁺	0.05	0.09	0.132	0.11	0.06
Ca ²⁺	57.0	158	147	99	110
DOC	1.50	49.5	38.5	11.0	6.76

Representative analysis of groundwater down-gradient of oil pool along plume axis. Elemental analyses were done by direct current plasma emission and results are reported in mg l^{-1} . Alkalinity was determined in the field by electrometric titration and reported as mg l^{-1} bicarbonate ion. pH and temperature were analysed using a flow-through cell¹⁵. Dissolved organic carbon was measured by the wet-oxidation method¹⁴, and is expressed as mg l^{-1} carbon.

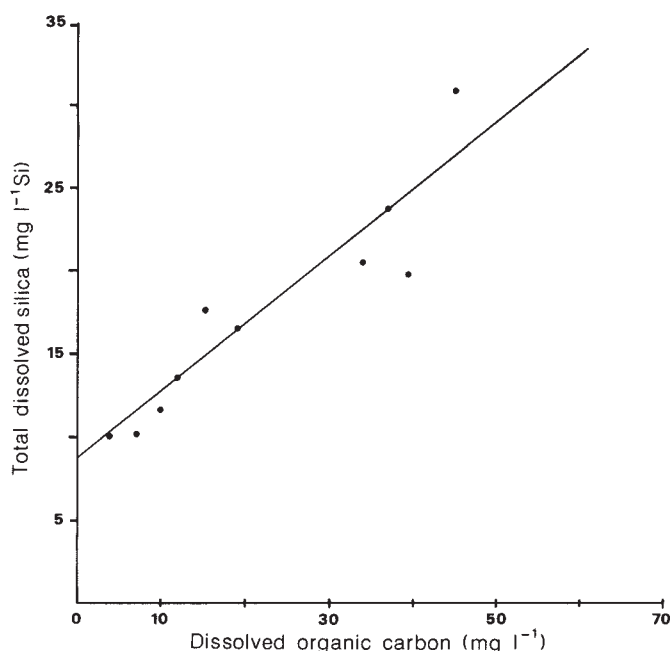


Fig. 3 Comparison of dissolved silica and dissolved organic carbon concentrations along the contaminant plume axis. Correlation coefficient $r = 0.83$.

a sediment of such recent origin (6,000 years) and in waters of neutral pH. Etch pits are not observed on grains that have been physically coated with oil (Fig. 2b), which precludes the possibility of a direct oil-solid interaction.

The etching of quartz is occurring in ground water apparently supersaturated with dissolved silica. The etching is greatest where concentrations of DOC are highest. Etching decreases with decreasing DOC. Etch features are not found on grains from uncontaminated wells. Concentrations of dissolved silica are strongly correlated with dissolved organic carbon concentration (Fig. 3).

Three models can be suggested to explain our observations. First, quartz may be dissolving because of an increased pH in the pore water micro-environment. However, detailed measurements of pH clearly show that pH decreases in the region of greatest etching, and reaches a value nearly 3 units below the pK_a of silicic acid. Second, a change in groundwater temperature may increase the solubility of quartz. Measurements of temperature, however, show no significant change along the contaminant plume. A third model is an organic-silica interaction.

We propose that silica is being complexed by organic compounds produced by the microbial degradation of crude petroleum. Microbial degradation, active at the site¹⁶, is producing a variety of soluble organic by-products. The major soluble by-products of the microbial oxidation of petroleum in water have been shown to be a variety of aliphatic and aromatic acids as well as hydroxy- and keto-acids¹⁸⁻²⁰. High concentrations of organic acids also occur in oil-field brine waters²¹. A variety of organic acids have been detected in the contaminated ground water at the study site, and may constitute a major portion of the dissolved organic carbon (M. J. Baedecker, personal communication).

Organic acids have been shown to complex with many metals in water^{22,23}. We propose that these acids, or possibly other organic compounds, may also be complexing dissolved silica, lowering the activity of silicic acid and accelerating the observed dissolution of quartz and aluminosilicate minerals at near neutral pH.

Complexes between organic acids and silica may contribute to the mobilization and transport of silica in both contaminated

and uncontaminated environments. For example, two mechanisms commonly used to explain the source of silica cement in sandstones are the dissolution of quartz grains in pore water with high pH, or from pressure dissolution during diagenesis². However, organic acids formed by the biodegradation of marine or freshwater organic material could complex, mobilize and transport silica during diagenesis at neutral pH and near-surface environmental conditions. A silica-organic acid complex may also help explain the rapid weathering of silicate mineral grains in environments of high organic content, such as soil zones and oil field sandstones.

We thank Dr M. J. Baedecker, M. Melcer, and the staff of the N. C. Brown Center for Ultrastructure Research for their assistance. This study was supported by the Toxic Waste Research program of the US Geological Survey, and the Graduate School of Syracuse University.

Received 23 September 1986; accepted 8 January 1987.

1. Iler, R. K. *Chemistry of Silica* (Wiley-Interscience, New York, 1979).
2. Blatt, H., Middleton, G. & Murray, R. *Origin of Sedimentary Rocks* (Prentice-Hall, Englewood Cliffs, New Jersey, 1972).
3. Chesworth, W. & Macias-Vasquez, F. *Am. J. Sci.* **285**, 128-146 (1985).
4. Van Lier, J. A., De Bruyn, P. L. & Overbeek, J. Th. G. *J. phys. Chem.* **64**, 1675-1682 (1960).
5. Fournier, R. V. & Potter, R. W. *Geochim. cosmochim. Acta* **10**, 1969-1973 (1982).
6. Livingston, D. A. *Proc. Pap. U.S. geol. Surv.* **440-G** (1963).
7. Chou, L. & Wollast, R. *Am. J. Sci.* **285**, 969-993 (1985).
8. Rimstidt, F. D. & Barnes, H. L. *Geochim. cosmochim. Acta* **44**, 1683-1699 (1980).
9. Ingri, N. in *Biochemistry of Silicon and Related Problems* (eds Bendz, G. & Lindqvist, I.) 3-52 (Plenum, New York, 1977).
10. Iler, R. K. in *Biochemistry of Silicon and Related Problems* (eds Bendz, G. & Lindqvist, I.) 53-75 (Plenum, New York, 1977).
11. Hult, M. F. *U.S. geol. Surv. Water-Res. Inv. Rep.* **84-4188** (1984).
12. Folk, R. L. *J. sed. Petrology* **25**, 297-301 (1955).
13. Myshlyayeva, L. V. & Krasnoshchekov, V. V. *Analytical Chemistry of Silicon* (Wiley, New York, 1974).
14. Baedecker, M. J. *2nd Toxic Waste Tech. Meeting*, Cape Cod, October 1985 (US Geol. Surv., in the press).
15. Wood, W. W. in *Techniques of Water-Resources Investigations of the U.S. Geological Survey*. Book 1, Chap. D2 (US Government Printing Office, Washington DC, 1981).
16. Chang, F., Hult, M., Hoben, N. & Brand, D. *2nd Toxic Waste Tech. Meeting* Cape Cod, October 1985 (US Geol. Surv., in the press).
17. Krinsley, D. H. & Doornkamp, J. C. *Atlas of Quartz Sand Surface Textures* (Cambridge University Press, 1973).
18. Norendova, I. K., Arkhangel'skaya, R. A. & Tarrasova, T. G. *Geokhimiya* **3**, 408-414 (1978).
19. Atlas, R. M. *Petroleum Microbiology* (MacMillan, New York, 1984).
20. Liu, D. L., Dutka, B. J. *J. WPCF* **45**, 232-239 (1973).
21. Surdam, R. C., Boese, S. W. & Crossey, L. J. in *Clastic Diagenesis* (eds McDonald, D. A., Surdam, R. C.) 127-150 (American Association of Petroleum Geologists, Tulsa, 1984).
22. Zutic, V. & Stumm, W. *Geochim. cosmochim. Acta* **48**, 1493-1503 (1984).
23. Huang, W. H. & Keller, W. D. *Am. Miner.* **55**, 2076-2094 (1970).

Selenium in the tributaries of the Orinoco in Venezuela

H. S. Yee, C. I. Measures & J. M. Edmond

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, E34-201, Cambridge, Massachusetts 02139, USA

The drainage basins of large pristine rivers are valuable natural laboratories for the study of trace element geochemistry. Their relatively unperturbed nature provides the opportunity to establish natural concentration ranges thus giving a background against which anthropogenically impacted environments may be understood. The wide range of geological provinces present in large basins also provides an opportunity to identify the underlying geochemical parameters which determine the sources and transport of the elements in the hydrosphere.

We have sampled the Orinoco river and its tributaries in Venezuela and have found total selenium concentrations ranging from 60 pM to 5,163 pM while selenium IV varies from <20 pM to 1,690 pM. The relatively low selenium concentrations in the Guyana Shield on the right bank of the river are dominated by rain inputs. The major input of selenium to the river system comes from the alkaline weathering of shale outcrops in the

Table 1 The data set from sampling in the Orinoco and its tributaries

Name	Date	Selenium Total	IV	Name	Date	Selenium Total	IV		
Caroni basin				Apure basin					
Caroni	6/82	255	<20	Apure	6/82	2,360	352 467		
	8/82	250			8/82	1,076			
	12/82	243			11/82	2,337			
	3/83	205			3/83	2,930			
				10/84	1,378				
Aro basin				6/85	2,430				
Aro	12/82	258	<20	Flood plain	10/84	788	<200		
	3/83	274			10/84	556			
Mapire	12/82	277		Bocono	10/84	1,402			
					Guanare	10/84		3,752	320
Cuchivero-Caura basin				Portuguesa	10/84	2,610		<200	
Cuchivero	6/82	267	<20	Morador	10/84	5,163	590		
	12/82	253		Acarigua	10/84	4,862	1,690		
	3/83	291		S. Domingo	10/84	497			
	10/84	244		Flood plain	10/84	637			
Caura	6/82	216		Canagua	10/84	223			
	8/82	298	Flood plain	10/84	215				
	12/82	207	Suripa	10/84	633				
	3/83	201	Caparo	10/84	394				
	10/84	253	Dorados	10/84	495				
Tortuga-Samariapo basins				Tururu	10/84	334			
Tortuga	11/82	185	<20	Uribante	10/84	703			
	11/82	93		Flood plain	10/84	1,042			
Suapure	3/83	179		Lower left bank					
	6/82	115		Manapire	12/82	957	199		
Parguaza	11/82	60		<20	Cabruta	11/82	438		
	3/83	92	Main channel at:						
	10/84	105	C. Bolivar		6/82	448			
	Cataniapo	11/82	94		8/82	412			
Paria Grande	11/82	164	12/82		527				
Samariapo	11/82	159	3/83	634	50				
Sipapo basin				P. Ordaz	6/82	435			
Sipapo	11/82	135	<20	8/82	527				
	3/83	153		12/82	509	101			
Cuao	3/83	188		Musinacio	12/82	611	107		
					3/83	619			
Ventuari basin				Cabruta	11/82	742	164		
Ventuari	3/83	76	<20	Capanaparo	3/83	479			
				Parguaza	6/82	280			
Atabapo	6/82	181							
11/82	245	11/82		250					
3/83	160	Apure		6/82	357				
10/84	202	Caura	6/82	422	<20				
Inirida	11/82	184	Atabapo	11/82	215				
	3/83	110							
Middle left bank									
Guaviare	11/82	280	<20						
	3/83	228							
Meta	6/82	759							
	8/82	660							
	11/82	892							
	10/84	798							
Cinaruco	11/82	163	<20						
Capanaparo	11/82	333							
	3/83	556							
Arauca	11/82	694		34					

Detection limits are based on volumes available for analyses and are normally around 20 pM for a 100 ml sample, 200 pM for a 10 ml sample.

foothills of the Andes on the lower left bank. The coincidence of this with an area in Venezuela where elevated selenium concentrations in crops are observed, suggests that relatively simple river surveys would highlight regions where weathering release of the element may be high enough to result in deleterious bio-accumulation.

All samples were collected using standard clean oceano-

graphic techniques and were filtered (usually in the field) through acid-cleaned 0.4 µm Nuclepore filters before acidification with 3×redistilled 6M HCl (1 ml per 250 ml). Determinations were made on volumes ranging from 10 to 100 ml using the technique of Measures and Burton¹ modified by the addition of 10 µmol sodium nitrate to the aliquots prepared for total selenium analyses to enhance the reproducibility

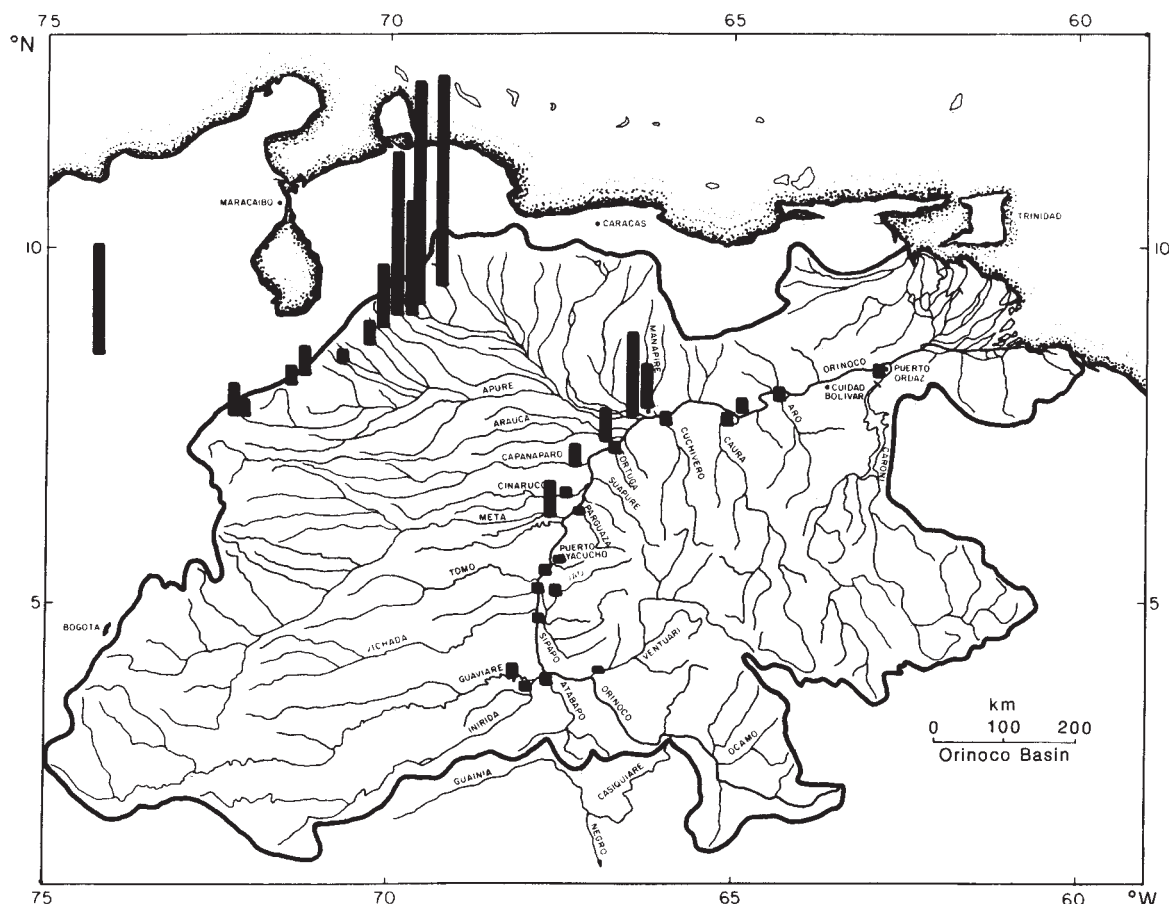


Fig. 1 Distribution of total selenium in selected tributaries of the Orinoco watershed. The isolated bar represents a concentration of 2,500 pM.

of recovery in fresh waters. The technique determines Se IV (selenite) directly, total selenium is determined as selenite after ultraviolet oxidation, a process which reduces Se VI (selenate) and which presumably oxidizes any reduced ($-II/0$) selenium present in the samples. The precision of the technique is $\pm 6\%$ for total Se and $\pm 3\%$ for Se IV; the limit of detection is 20 pM for a 100 ml volume. Occasional replication was used to assure both the overall consistency of the data set and its conformation to the accepted precision of the technique.

The geology of the Orinoco basin may be crudely divided into three regions. The right bank (looking downstream) of the river in the headwaters drains a deeply weathered Precambrian lowland terrain; further downstream, right bank tributaries run off the elevated Guyana Shield. The left bank of the river is fed by tributaries whose headwaters rise in the shallow marine and brackish water lithologies of the Colombian and Venezuelan Andes and flow across the Llanos, an area of recent fluvio-lacustrine sediments and shales, before joining the Orinoco. The major sedimentary and solute input to the river comes from these left bank tributaries. A more detailed geological description will be presented elsewhere (J.M.E., in preparation). Previous work on other individual rivers has been reviewed recently by Cutter².

The range of values present within the data set (Table 1) is a reflection of the variations in the weathering rates and the nature of the geological substrate throughout the watershed overprinting the external sources of the element to the system (Fig. 1). In the headwaters and the tributaries deriving from the right bank Se levels are low, 60 pM–200 M. Reported concentrations of selenium in the types of acid rocks which comprise this region (granites, granodiorites) are in the range of 0.005 to 0.05 p.p.m. by weight^{3–5}. Congruous dissolution of such material would yield Se/Si molar ratios of about $5\text{--}50 \times 10^{-9}$. The Se/Si ratios

observed in these rivers are uncorrelated and more typically of the order of 1×10^{-6} . This, coupled with the very low dissolved load in these rivers (total cationic charge of the order $100 \mu\text{eq}^{-1}$) indicates that weathering of the host rock is an unlikely dominant source of the element. The relative constancy of the selenium concentrations both areally and temporally suggest that the values seen in these rivers in fact represent pristine rain. This is strongly supported by the data from the Atabapo (160 pM–257 pM) which, on the basis of analyses for the major constituents (Ca, SO_4 , Si, Mg), appears to have virtually no input from weathering. Three rain samples, collected at the Suapure confluence, Cabruta and C. Bolívar, yielded values of <170, 494 and 721 pM respectively. Although it is not possible to make any detailed interpretation of such a restricted data set, it should be noted that these values are similar to the volume-weighted average observed by Cutter and Church⁶ in Bermuda rain (380 pM). The occurrence of the highest value in the most downstream and urban part of the basin (C. Bolívar) may be evidence, albeit limited, of an anthropogenic input.

Although the left bank streams generally have higher selenium values, it is the Apure, and specifically four of its tributaries, that are the major sources of total selenium and selenite to the Orinoco. The headwaters of these four rivers, Acarigua, Morador, Guanare and Portuguesa, all arise in mountainous terrain dominated by outcrops of Upper Cretaceous sediments. They were sampled in the foothills of the Andes. The strong correlation in these and the other Orinoco samples between total selenium and calcium ($r = 0.90$, $n = 87$) and sulphate ($r = 0.84$) would, at first sight, appear to imply an evaporite-carbonate dominated rock source for the high selenium concentrations. However the Ca/Se molar ratio of 3.7×10^{-6} implies a whole-rock Se concentration of 2.9 p.p.m. by weight, much higher than the 0.08 p.p.m. reported by Turekian and Wedepohl⁷

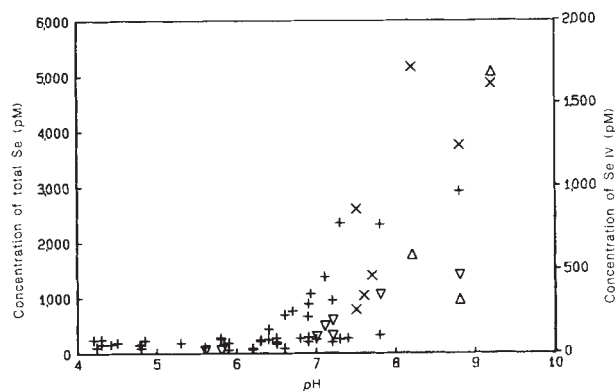


Fig. 2 Concentration of Se IV (∇) and total Se (+) as a function of pH. Where pH values were unavailable they have been interpolated from the alkalinity of the samples and its relationship to pH observed in other samples of the same region. These are designated (Δ) Se IV and ($\times$) total Se.

for limestones. The molar selenium/sulphate ratio of 7.5×10^{-6} is also hard to reconcile with marine evaporites in which a ratio similar to that of surface sea water (1.8×10^{-8}) would be expected. The occurrence of the highest values in samples collected before the rivers cross the Llanos precludes the reworking of fluvial or lacustrine sediments as a source. This conclusion is reinforced by data from the Capanaparo and Cinaruco which, as the only left bank rivers whose drainage is entirely within the Llanos, are observed to contain relatively low levels of selenium. Weathering of Upper Cretaceous calcareous black shales would appear to be the most likely explanation of the elevated levels of selenium in this region. Black shales in general are known to contain high levels of selenium⁸ probably in a reduced form (0 or -II). Data from the Julia Creek⁹, Green river, and Bear Paw shales (K. L. Von Damm, personal communication) show pyritic Se/S levels in the range 10^{-3} to 10^{-5} . Unfortunately no other measured tracer in the Orinoco data set is able to quantify the shale weathering signal.

The relationship between oxidation states of selenium and pH in this data set (Fig. 2) imply that, under the conditions prevalent in the drainage, mobilization of reduced Se is proceeding at pH above 7; this is in broad agreement with the work of Howard¹⁰. The technique used will include any -II and 0 oxidation state selenium in with the +VI state, but it seems likely that the majority of the fraction, that is the difference between total selenium and selenite, is composed of selenate and not dissolved organic selenium. Support for this argument comes from the relatively low total selenium observed in the black, low pH, organic-rich rivers of the Shield. More significantly though, organic selenium levels are only some 10–15% of the total even in the highly contaminated agricultural drainage waters which feed Kesterson reservoir in California¹¹. The occurrence of selenate as the dominant species in the Orinoco samples would also imply that oxidation potentials in the Se VI stability field¹⁰ are reached in natural weathering systems. When this is coupled with shales that contain high levels of carbonates, such that weathering pHs above 7 are achieved, significant mobilization of selenium is observed.

The region of high fluvial selenium concentrations appears to coincide with an area of Venezuela reported to produce vegetation with potentially toxic levels of the element^{12,13}. It is well known that shales frequently weather to soils in which the bio-availability of the element is such that seleniferous vegetation is produced⁸. The fact that the dissolved selenium levels in rivers and groundwater appear to reflect this bio-availability suggests that a relatively simple river survey of a region could be a useful precursor in identifying areas prone to such problems.

Total selenium in the Amazon (645 pM), the Chang Jiang

(2,953 and 2,802 pM), and the Hudson (1,081 pM) fall within the range observed for the Orinoco. Selenium IV values were 155 (Amazon), 490 and <170 (Chiang Jiang) and <170 pM (Hudson). Using the Amazon concentration of 645 pM, the annual riverine flux of total selenium to the oceans is 2.3×10^7 moles. This would imply an oceanic residence time of some 150,000 yr. This value, although some seven times higher than that estimated by Goldberg *et al.*¹⁴ must be considered somewhat unconstrained as it is based only on river data with no provision made for hydrothermal or volcanic inputs or recycled marine emissions which may be important.

We thank Ing. Abel Mejia of Proyecto Orinoco-Apure (Ministerio del Ambiente y de los Recursos Naturales Renovables) for his enthusiasm and support of our work in the Orinoco. Barry Grant and Bob Stallard collected many of the samples. We thank Dennis Burton, Ken Bruland, and Philip Froelich aided by Greg Cutter for their thoughtful and constructive reviews. H.S.Y. thanks B. Giletti for his advice and guidance. This work was supported by the Earth Surface Processes Section of the NSF.

Received 29 September 1986; accepted 13 January 1987.

1. Measures, C. I. & Burton, J. D. *Analytica chim. Acta* **120**, 177–186 (1980).
2. Cutter, G. A. in *Occurrence and Distribution of Selenium* (ed. Ilnatt, M.) (CRC Press, in the press).
3. Leutwein, F. in *Handbook of Geochemistry II-3* (ed. Wedepohl, K. H.) 2 (Springer, New York, 1978).
4. Kolmer, H. W. & Raptis, S. E. *Geostand. Newsl.* **7**, 315–318 (1983).
5. Imai, N., Terashima, S. & Ando, A. *Geostand. Newsl.* **8**, 39–41 (1984).
6. Cutter, G. A. & Church, T. M. *Nature* **322**, 720–722 (1986).
7. Turekian, K. K. & Wedepohl, K. H. *Bull. geol. Soc. Am.* **72**, 175–192 (1961).
8. Lakin, H. W. in *Trace Elements in the Environment* (ed. Kothny, W. L.) (American Chemical Society, Washington DC, 1971).
9. Patterson, J. H., Ramsden, A. R., Dale, L. S. & Fardy, J. J. *Chem. Geol.* **55**, 1–16 (1986).
10. Howard, J. H. *Geochim. cosmochim. Acta* **41**, 1665–1678 (1977).
11. Cooke, T. & Bruland, K. W. *Envir. Sci. Technol.* (submitted).
12. Jaffe, W. G., Chavez, J. F. & Mondragon, M. C. *Archos. lat.-am. Nutr.* **17**, 59–58 (1967).
13. Jaffe, W. G., Ruphael, M., Mondragon, M. C. & Cuevas, M. A. *Archos. lat.-am.* **22**, 595–612 (1972).
14. Goldberg, E. D., Broecker, W. S., Gross, M. G. & Turekian, K. K. in *Radioactivity in the marine environment* 137–146 (Natl. Acad. Sci. U.S.A., Washington DC, 1971).

An analogue approach to the travelling salesman problem using an elastic net method

Richard Durbin* & David Willshaw†

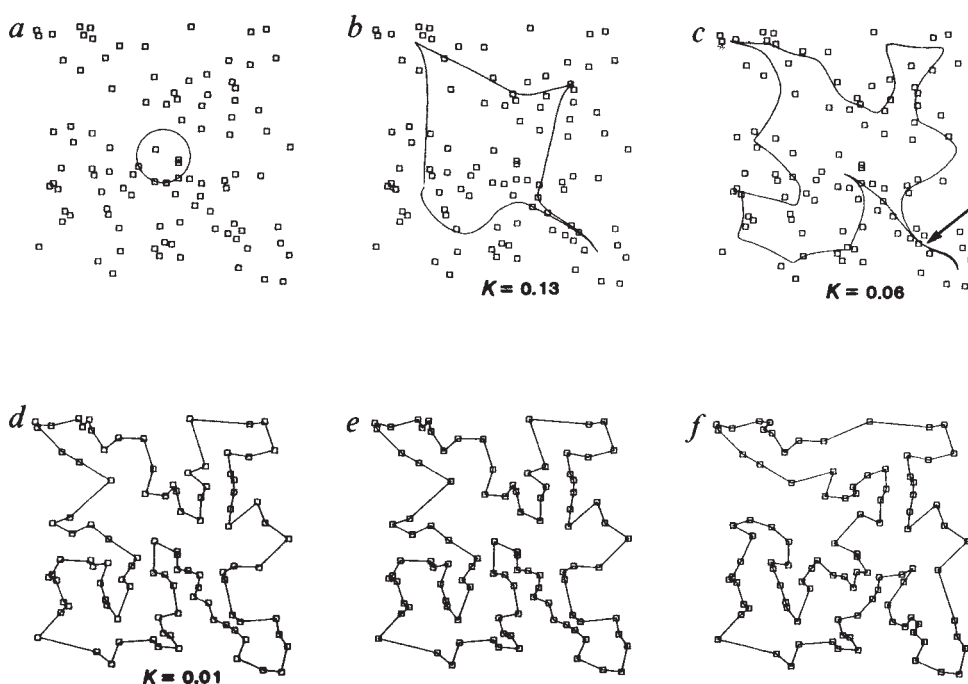
* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH and King's College Research Centre, Cambridge CB2 1ST, UK

† Department of Zoology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

The travelling salesman problem¹ is a classical problem in the field of combinatorial optimization, concerned with efficient methods for maximizing or minimizing a function of many independent variables. Given the positions of N cities, which in the simplest case lie in the plane, what is the shortest closed tour in which each city can be visited once? We describe how a parallel analogue algorithm, derived from a formal model^{2–3} for the establishment of topographically ordered projections in the brain^{4–10}, can be applied to the travelling salesman problem^{1,11,12}. Using an iterative procedure, a circular closed path is gradually elongated non-uniformly until it eventually passes sufficiently near to all the cities to define a tour. This produces shorter tour lengths than another recent parallel analogue algorithm¹³, scales well with the size of the problem, and is naturally extendable to a large class of optimization problems involving topographic mappings between geometrical structures¹⁴.

Although easy to state, the travelling salesman problem (TSP)

Fig. 1 Example of the progress of the elastic net method for 100 cities randomly distributed in the unit square. *a*, The initial path. To break symmetry, this was set up to be a ring of radius 0.1 about the centroid of the cities. A starting configuration with each point on the path placed at a small, randomly chosen distance away from the centroid (see ref. 13) was found to be adequate as long as a large number of iterations at a high value of K were allowed. *b*, *c* and *d*, Paths generated as the value of K was gradually lowered. *e*, The tour of length 7.78 deduced from the final configuration shown in *d*. *f*, The shortest tour so far found by us by any method (simulated annealing^{15,16} with eight million trials; see Table 1 for details). This tour had a length of 7.70. The parameter values used for the computations described in this paper which used the elastic net method were: $M = 2.5N$, $\alpha = 0.2$, $\beta = 2.0$; the initial value of K was 0.2, and was reduced by 1% every n iterations to a final value in the range 0.01–0.02. In the calculation shown here, K was reduced to 0.01 in 7,000 iterations by taking $n = 25$. The rate at which K can be reduced is limited by situations such as that shown by the arrow in *c*, where the system must be sufficiently relaxed to prevent the occurrence of crossovers, which would increase the tour length. The function ϕ had the form $\phi(d, K) = \exp(-d^2/2K^2)$. In this case an energy function E can be defined as $E = -\alpha K \sum_i \ln \sum_j \phi(|x_i - y_j|, K) + \beta \sum_j |y_{j+1} - y_j|^2$. This has the property that $\Delta y_j = -K \partial E / \partial y_j$, which means that any change in y_j according to equation (1) results in a reduction in the value of E and, because E is bounded below, local minima of E will eventually be reached. In the limit where K tends to zero, for E to remain bounded the limiting path must pass through all the cities (for every i there must be a j such that $|x_i - y_j|$ tends to zero). As M gets large, the second term in the expression for E is then minimized by spacing the M path points equally around the path at a distance D/M apart, where D is the path length. This term then takes the value D^2/M , which is minimized by minimizing the tour length D .



is hard to solve; the number of possible tours increases exponentially with the number of cities. It is representative of the important class of optimization problems^{1,12}, called NP-complete¹¹, which are all interconvertible, but the computing effort needed to solve them increases faster than any power of N . Because perfect solutions are unattainable for problems of any appreciable size, there has been interest in heuristic methods that find near-optimal solutions in a reasonable time¹.

Hopfield and Tank¹³ have recently discussed how problems such as the TSP can be represented in the dynamical behaviour of a network of simple model neurons. The activity of the network evolves into a stable configuration which represents a good solution to the problem. This procedure was shown to be equivalent to gradient descent of an energy function which contains the tour length as a major term. For the 30-city problem discussed, the best tour found was only 19% longer than the shortest known tour. A related analogue algorithm (G. Tesauro, personal communication) obtains shorter tours for 30 cities, but does not, in general, produce valid solutions for larger problems.

Our approach is more geometrical. A tour can be viewed as a mapping from a circle to the plane so that each city in the plane is mapped to by some point on the circle. We consider mappings from a circular path of points to the plane in which neighbouring points on the circle are mapped as close as possible on the plane. This is a special case of the general problem of best preserving neighbourhood relationships when mapping between different geometrical spaces. Our algorithm was developed from the tea-trade model^{2,3} for the establishment of topographically ordered, neighbour-preserving projections between neural structures with matching geometries^{4–10}, such as the projection of the vertebrate retina onto the optic tectum^{5,6}.

The algorithm is a procedure for the successive recalculation of the positions of a number of points in the plane in which the

cities lie. The points describe a closed path which is initially a small circle centred on the centre of the distribution of cities and is gradually elongated non-uniformly to pass eventually near all the cities and thus define a tour around them (Fig. 1). Each point on the path moves under the influence of two types of force. The first moves it towards those cities to which it is nearest; the second pulls it towards its neighbours on the path, acting to minimize the total path length. By this means, each city becomes associated with a particular section of the path. The tightness of the association is determined by how the force contributed from a city depends on distance, and the nature of this dependence changes as the algorithm progresses. Initially all cities have a roughly equal influence on each point on the path. Subsequently, longer distances become less favoured, and each city gradually becomes more specific for the points on the path closest to it. This gradual increase of specificity (informally equivalent to the lowering of the 'temperature' in optimal simulated annealing^{15,16}) is controlled by a reduction of the length parameter K . We call this algorithm the elastic net method because of the way in which the initial path is gradually deformed to produce the final tour.

Let the coordinates of the position of a typical city i be denoted by the vector x_i and those of a typical point j on the path by y_j . Then the rule for the change Δy_j in the coordinates y_j of point j at each iteration is

$$\Delta y_j = \alpha \sum_i w_{ij}(x_i - y_j) + \beta K(y_{j+1} - 2y_j + y_{j-1}) \quad (1)$$

where the constants α and β determine the relative strengths of the forces from the cities and the forces from its neighbours on the path. The coefficient w_{ij} specifies the influence of city i on path point j , and can be thought of as the strength of the connection between i and j . This is a function of the distance $|x_i - y_j|$ between i and j and of the length parameter K . It is

Table 1 Tour lengths obtained from comparative studies on five sets of 50 randomly positioned cities

City set	Elastic net method*	Simulated annealing†	All algorithms‡
1	5.98	5.88	5.84, 3
2	6.03	6.01	5.99, 3
3	5.74 [5.70]	5.65	5.57, 6
4	5.90 [5.86]	5.81	5.70, 4
5	6.49	6.33	6.17, 1

* For each set of cities, the algorithm was run for 1,250 iterations, with $n = 5$. For sets 3 and 4, a slower schedule for reducing the value of K resulted in shorter tours, shown in brackets.

† At each trial, three cuts in the tour were made at random and rearrangements which produced shorter tours were accepted; in the absence of a shorter tour, the rearrangement which lengthened the tour least was accepted with probability $\exp(-\Delta D/T)^{15,16}$, where ΔD is the increase in tour length and the parameter T corresponds to the parameter K of the elastic net method. T was reduced from 0.5 to 0.003 (5×10^5 steps, $n = 1,000$), which took 30% more computer time than the elastic net method. Each figure is the average of 5 simulations.

‡ Tour optimization was also attempted by systematic trial of all possible three-cut rearrangements and accepting only those rearrangements which shortened the tour, the procedure being repeated from many different randomly generated initial configurations. This strategy has been shown to be efficient in the use of time for this size of problem²⁹. Each pair of figures in the final column represents respectively the shortest tour obtained from any of these methods and the number of times this tour was obtained.

normalized so that the total influence of each city is equal.

$$w_{ij} = \phi(|x_i - y_j|, K) / \sum_k \phi(|x_i - y_k|, K) \quad (2)$$

where $\phi(d, K)$ is a positive, bounded decreasing function of d that approaches zero for $d > K$.

Taking $\phi(d, K)$ as the gaussian function $\exp(-d^2/2K^2)$, an energy function can be defined which has the property that successive changes in the path calculated according to equation (1) lead to a stable state which represents a local minimum of energy (see Fig. 1 legend). In the limit where K tends to zero and the number of points M on the path tends to infinity, the global minima of the energy function are optimal solutions of the TSP. This suggests that even when these limits are not achieved, good tours will be obtained.

The elastic net method was applied to the distribution of 30 cities used by Hopfield and Tank¹³. It generated the shortest known tour¹⁷ in 1,000 iterations; Hopfield and Tank's shortest tour was 19% longer. A comparison of our algorithm with more conventional algorithms is shown in Fig. 1 and Table 1. The length of the best tour around 100 cities distributed at random in the unit square was 7.78, which is within 1% of the best tour length (7.70) we have found by any method (Fig. 1). When applied to 5 sets of 50 randomly distributed cities, the best tour length was within $3.0 \pm 1.4\%$ (mean $\pm$ s.d.) of the best tour length we have found by any method, and within $1.5 \pm 0.7\%$ of the average length we found by simulated annealing^{15,16} in approximately the same amount of computing time (Table 1).

The elastic net method operates by integrating a set of simultaneous first-order difference equations, an essentially parallel operation, and it therefore naturally lends itself to implementation in parallel hardware. Even when implemented on conventional computers, its complexity scales well with the number of cities N . It produces shorter tours than the parallel algorithm of Hopfield and Tank¹³ and it is not as sensitive to the choice of parameter values. The main computational cost is in calculating the distances at each iteration and is therefore proportional to the product of the number of iterations required with the number of significant connections (connections with near-zero strength can be ignored). The number of connections is of the order N^2 (given that the number of points on the path is of

order N), but this could be reduced by starting with a small number of path points and gradually interpolating additional points in the path to keep a constant average number of significant connections from each city. The number of connections required by the algorithm of Hopfield and Tank¹³ is of the order of N^3 , although these are fixed in value. The number of iterations required seems to increase more slowly than a linear function of N , and this number could probably be significantly decreased by optimizing the schedule for decreasing K . Because an energy function can be defined, a more sophisticated approach to minimization¹⁸ might also reduce the computational time required.

Our algorithm is fundamentally geometrical. It can be extended to the more general TSP with cities in any euclidean space, but not to the case where an arbitrary matrix of distances is given. However, it is applicable to a wide range of geometrical optimization problems. For instance, we can alter the dimension and topology of either space (by, for example, replacing the ring by any other graph), and also remove the condition that points in the 'ring' space are to be evenly spaced (by varying the strength of the force between neighbours). The algorithm therefore provides a general method for matching a set of arbitrarily connected points to a second set of points located in a geometrical space of any dimensionality. It may thus be applicable to a variety of matching and representational problems^{1,16,19}, such as those that arise in early visual processing^{20,21}.

Our interest in this problem originated in an interest in mechanisms^{2,3} for the establishment of ordered neural projections between structures of similar geometry. There are other neural projections which involve topographical mappings between spaces of different dimensionality²²⁻²⁴, for which several formal models have been proposed²⁵⁻²⁸. The elastic-net approach suggests a new way of looking at these phenomena and we are currently investigating how simple parallel algorithms of the type discussed here could give rise to the projection patterns observed.

We thank the participants of the Connectionist Summer School, held at Carnegie-Mellon University, USA in May 1986, for their interest and their help, particularly Dr G. Tesauro. We also thank Dr A. Yuille for the suggestion of a form of the function $\phi(d, K)$ which permits the construction of an energy function. Dr D. W. Tank kindly supplied the coordinates of the 30-city problem discussed.

Received 21 November 1986; accepted 18 February 1987.

- Lawler, E. L., Lenstra, J. K., Rinooy Khan, A. H. G. & Shmoys, D. B. (eds) *The Traveling Salesman Problem* (Wiley, New York, 1985).
- Von der Malsburg, Ch. & Willshaw, D. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5176-5178 (1977).
- Willshaw, D. J. & Von der Malsburg, Ch. *Phil. Trans. R. Soc. B* **287**, 203-243 (1979).
- Talbot, S. A. & Marshall, W. H. *Am. J. Ophthalmol.* **24**, 1255-1264 (1941).
- Gaze, R. M. *The Formation of Nerve Connections* (Academic, London, 1970).
- Cowan, W. M. & Hunt, R. K. in *The Molecular Basis of Neural Development* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 389-428 (Wiley, New York, 1986).
- Rose, J. E., Galambos, R. & Hughes, J. R. *Bull. Johns Hopkins Hosp.* **104**, 211-251 (1959).
- Knudsen, E. I. & Konishi, M. *Science* **200**, 795-797 (1978).
- Drager, U. C. & Hubel, D. H. *J. Neurophysiol.* **38**, 690-713 (1975).
- Wurtz, R. H. & Albano, J. E. *Rev. Neurosci.* **3**, 189-226 (1980).
- Garey, M. R. & Johnson, D. S. *Computers and Intractability* (Freeman, San Francisco, 1979).
- Papadimitriou, C. H. *Theor. Comput. Sci.* **4**, 237-244 (1977).
- Hopfield, J. J. & Tank, D. W. *Biol. Cybern.* **5**, 141-152 (1985).
- Mitchison, G. J. & Durbin, R. *SIAM J. alg. disc. Meth.* **7**, 571-582 (1986).
- Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. *Science* **220**, 671-680 (1983).
- Kirkpatrick, S. *J. statist. Phys.* **34**, 975-986 (1984).
- Lin, S. & Kernighan, B. W. *Oper. Res.* **21**, 498-516 (1973).
- Leuenberger, D. G. *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, Reading, Massachusetts, 1984).
- Christofides, N. in *The Traveling Salesman Problem* (eds Lawler, E. L., Lenstra, J. K., Rinooy Khan, A. H. G. & Shmoys, D. B.) 431-448 (Wiley, New York, 1985).
- Julesz, B. *Foundations of Cyclopean Vision* (University of Chicago, 1971).
- Marr, D. & Poggio, T. *Proc. R. Soc. B* **204**, 301-328 (1979).
- Hubel, D. H. & Wiesel, T. N. *Proc. R. Soc. B* **198**, 1-59 (1977).
- Law, M. I. & Constantine-Paton, M. *J. Neurosci.* **7**, 741-759 (1981).
- Fawcett, J. W. & Willshaw, D. J. *Nature* **296**, 350-352 (1982).
- Von der Malsburg, Ch. & Willshaw, D. J. *Expl Brain Res. (Suppl.)* **1**, 463-469 (1976).
- Von der Malsburg, Ch. *Biol. Cybern.* **32**, 49-62 (1979).
- Swindale, N. V. *Proc. R. Soc. B* **208**, 243-264 (1980).
- Kohonen, T. *Self-organisation and Associative Memory* (Springer, Berlin, 1984).
- Brady, R. M. *Nature* **317**, 804-806 (1985).

A fossil owl monkey from La Venta, Colombia

Takeshi Setoguchi* & Alfred L. Rosenberger†

* Primate Research Institute, Kyoto University, Inuyama City, Aichi 484, Japan

† Department of Anthropology, University of Illinois at Chicago, Chicago, Illinois 60680, USA

Knowledge of the evolutionary history of living New World anthropoids is limited by a relatively poor fossil record. The discovery in 1986 of a new fossil monkey from the middle Miocene deposits of La Venta, Colombia, 12–15 million years ago (Myr BP), is the first example of a living New World monkey genus appearing in Tertiary rocks. Including anatomical evidence of the dentition and facial skull, it provides an unambiguous link between a Neogene fossil and the owl monkey, *Aotus*, the only modern crepuscular-nocturnal anthropoid primate. This new form brings to three the number of La Venta fossil monkeys which preserve excellent dentitions sharing extensive similarities with modern genera. All of these species are potentially ancestral to their extant relatives. The La Ventan *Aotus* is additional support for the idea that the modern platyrrhine radiation includes long-lived genera or generic lineages, some of which may be traceable to the early Miocene, 20 Myr BP.

The broad diversity of the living platyrrhine monkeys of Central and South America has led to the belief that they underwent a very early adaptive radiation¹. Although the platyrrhine fossil record is still too scarce to evaluate this hypothesis properly, recent reviews of the material² tend to support the idea, bolstered by the identification among the early and middle Miocene species of lineages pertaining to modern genera^{3,4}. We report here the discovery of a new species which confirms the early origins of one extant form with some clarity, for it appears to be taxonomically indistinguishable from the owl monkey, *Aotus*, at the generic level.

Order Primates Linnaeus, 1758

Suborder Haplorhini Pocock, 1918

Infraorder Platyrrhini E. Geoffroy, 1812

Family Atelidae Gray, 1825

Subfamily Pitheciinae Mivart, 1865

Aotus Illiger, 1811

Aotus dindensis, sp. nov.

Type specimen. IGM-KU (Instituto Nacional de Investigaciones Geológico-Mineras [INGEOMINAS]-Kyoto University) 8601, a left hemimandible with right I₁-left M₃ and a left maxillary fragment preserving roots of P³-M² and the lingual half of M³, all of the same individual.

Hypodigm. Type specimen only.

Locality. Locality 9-86-A in the El Dinde area, probably within the Monkey Unit of the Honda Formation⁵, in the Tatacoa desert, Huila Department, Republic of Colombia.

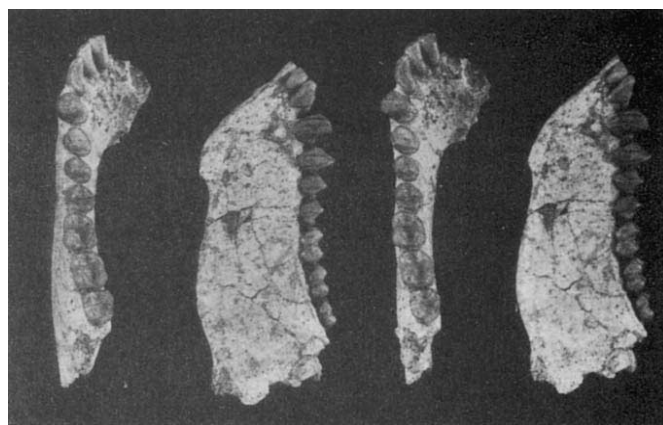


Fig. 1 Stereo pairs of *Aotus dindensis*, n. sp., IGM-KU 8601 (x2). Left pair, occlusal view; Right pair, buccal view.

Age. Middle Miocene, Friasian Land Mammal Age, 12–15 Myr (ref. 6).

Etymology. After El Dinde, the area where the fossil was recovered during the 1986 INGEOMINAS-KU field season.

Diagnosis. Differs from living species of the genus *Aotus* in the following combination of characters: lower incisors smaller in absolute size and in relation to cheek teeth; P_{3,4} metaconids smaller and less distinct; premolar trigonids less elevated; P₄ entoconid undifferentiated and P_{3,4} talonids weakly developed; molar trigonids small with metaconid displaced distally, rather than broad and quadrate with transverse cusps; molar metaconids and its crests less prominent (see measurements in Table 1).

Description and discussion. So far as is known, *A. dindensis* shares with the living species of *Aotus* the same diagnostic suite of dental and mandibular features which distinguishes the latter from other platyrrhine genera (Figs 1–4). This includes, in combination: I_{1,2} with recumbent crowns; P₂ with buccolingually compressed, tall and pyramidal crown (in both sexes among the modern species); P_{2,3} with prow-like preprotocristid; P_{3,4} not transversely broadened; M_{1,2} with tall trigonids and faintly differentiated talonid cusps; anterior symphyseal surface slopes forward, is flattened and has moderate lower canine root bosses; inferior border of mandibular corpus sigmoidal, slightly inflected below molars and deepening posteriorly; lingual aspect with large digastric fossa. The only taxonomically distinguishing features, therefore, are those pertinent to the species diagnosis, most notably incisor proportions, and relatively minor aspects of premolar and molar shape.

The maxillary fragment preserves 10 mm of the dental arcade inferiorly and a portion of the orbital floor superiorly. Essentially, it comprises the compact bone of the maxillary process, securing the roots of the last four postcanines (Fig. 2) and preserving the P³ alveolus and the maxillary tuberosity behind M³. The transverse diameter of this fragment is 11 mm, from the facial surface laterally to its medial limit. The palatal surface preserves small portions of the palatine and the alisphenoid-

Table 1 Dimensions of *Aotus dindensis*, IGM-KU 8601

	I ₁	I ₂	C	P ₂	P ₃	P ₄	M ₁		M ₂		M ₃	
							Tr	Tl	Tr	Tl	Tr	Tl
L	1.70	1.81	2.80	2.85	2.60	2.80	3.52		3.73		3.03	
B	2.20	2.51	3.30	2.62	2.40	2.69	2.80	3.02	2.94	2.92	2.80	2.56
H	3.30	3.70										
Mandible height below M ₂ :		10.00										
Mandible thickness at M ₂ :		3.35										

L, mesiodistal; B, buccolingual; H, unworn height; Tr, trigonid; Tl, talonid. All dimensions in mm.

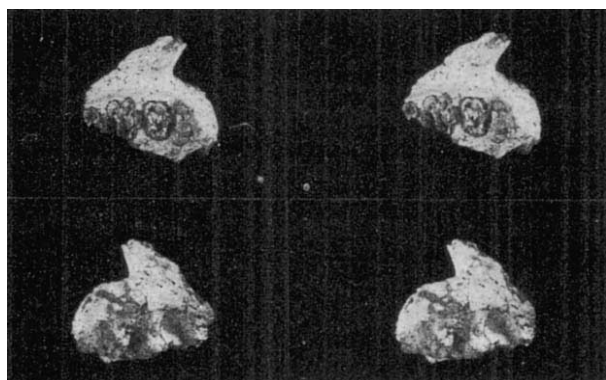


Fig. 2 Stereo pairs of *Aotus dindensis*, n. sp., IGM-KU 8601, maxillary fragment ($\times 2$). Top, occlusal view; bottom, ventral view.

palatine suture. The greatest dorsoventral thickness from orbital floor to M^1 alveolus is 3 mm. Thus the orbital floor was not elevated relative to the toothrow. Only a small anterior portion of the orbital floor appears to have been invaded by maxillary sinus, and there was probably none above the molars posteriorly.

The thick root of the zygomatic arch lies immediately above the alveolar plane near M^2 . Its posteroinferior aspect forms a deep concave incisure relative to the lateral surface of the maxillary tuberosity, which further extends for several millimeters posteriorly. This represents the inferior orbital fissure. The size and shape of this opening indicates that the fissure was large in its pristine state, comparable in all details to the anatomy of *Aotus* which is unique among living anthropoids^{1,7}. There are no scars or flattenings on the dorsal surface of the maxilla near the fissure which might indicate ossification or contact with the zygomatic, which forms the posterior seal of the orbit in other anthropoids. Anteriorly, a jagged edge of broken bone marks the position of the ascending lateral wall of the eye socket. This pattern makes it clear that, as in *Aotus*, the orbitotemporal fenestra of *A. dindensis* was large and the internal aspect of the orbit capacious, conforming to a relatively large eyeball.

Aotus dindensis is dentally more primitive than modern *Aotus*. The latter is unique among all living and fossil platyrrhines in exhibiting a scoop-like battery of relatively large lower incisors, distinctly absent in the fossil. Additionally, its lower cheek teeth support taller premolar metaconids and taller, more sharply crested trigonids. Whereas living *Aotus* and *A. dindensis* are probably identical in body size, these morphological contrasts suggest some adaptive differences. The fossil was perhaps less adept in processing foods by premolar and molar puncture-crushing and shearing, implying less insectivory. The smaller incisors indicate a lesser development of harvesting functions, either of fruit husks or insect-infected tree bark. Owl monkeys are mixed feeders, relying on high proportions of fruits and some insects and leaves⁸.

The cranial material, although very fragmentary, leaves little doubt that the orbits of *Aotus dindensis* were enlarged. The eyeballs of living *Aotus* are very large relative to body size. Their proportions compare with large-eyed nocturnal strepsirrhines⁹, and in both absolute and relative terms their eyes are considerably larger than in other ceboids¹⁰. The living *Aotus* is nocturnal throughout most of its geographical range, although its activity pattern involves more daytime and crepuscular hours among populations occupying nontropical southern latitudes¹¹. *Aotus dindensis* probably had a similar lifestyle, albeit with a less 'specialized' diet.

The ensemble of similarities shared by the modern *Aotus* and *A. dindensis* spans two disjunct functional complexes, the masticatory and visual systems. Thus *Aotus dindensis* is the first fossil platyrrhine to be phylogenetically linked with living owl monkeys on the basis of morphological and adaptive patterns of high taxonomic weight. Two Patagonian fossil forms are per-

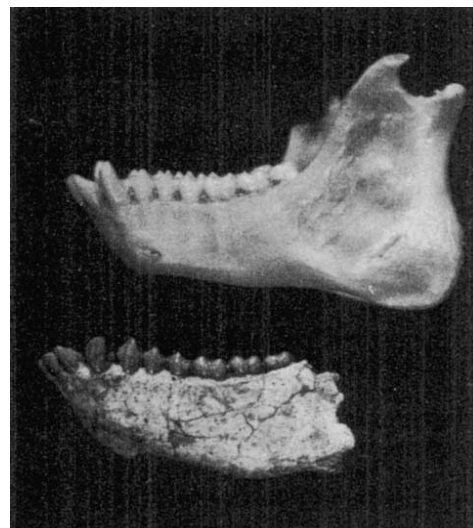


Fig. 3 Comparison of *Aotus trivirgatus* (top) and *A. dindensis* (below). Buccal view of mandibles ($\times 1.75$).

tinent here. The type specimen of the early Miocene *Tremacebus harringtoni* jointly shares with *Aotus* derived features of the orbits as well as the palate¹², but its dentition is poorly preserved and the molar pattern is difficult to reconstruct and interpret. A mandibular fragment recently allocated to *Tremacebus* is reportedly most similar to *Homunculus*¹³, however. The later early Miocene *Homunculus*, which is known from better material, does not resemble *Aotus dindensis* in occlusal morphology but is comparable in lateral mandibular profile. However, this pattern is generally primitive for pitheciines¹⁴, and the cranial anatomy of *Homunculus* is best interpreted either as evidence of closer affinities with *Callicebus*¹², or as primitive pitheciine morphology.

Two other La Venta primates known by excellent dentitions, *Neosaimiri fieldsi* and *Stirtonia tatacoensis*, also share extensive similarities with modern forms. They are closely related to *Saimiri* and to *Alouatta*, respectively⁴. Among the nonprimate terrestrial mammals, two species are congeneric with the extant murine or mouse opossum, *Marmosa*¹⁵, one of the edentates is thought to be directly ancestral to the living giant anteater, *Myrmecophaga*¹⁶, and several of the rodents are reportedly very similar and potentially ancestral to modern genera¹⁷. This association of enduring lineages at La Venta may have important consequences for macroevolutionary interpretations. It suggests that local or regional conditions during the Neogene may have promoted a widespread bathyphyly of the modern indigenous South American mammals, one not limited to the primates. On the other hand, La Venta's larger marsupial ungulates, litopter-

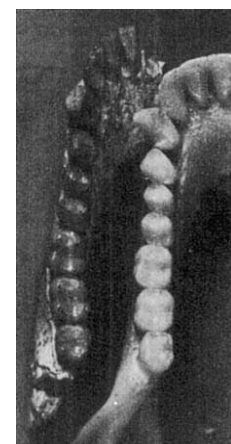


Fig. 4 Comparison of *Aotus trivirgatus* (right) and *A. dindensis* (left). Occlusal view ($\times 2.3$).

nans, notoungulates, astrapotheres and possibly condylarths left no descendants, just as they became extinct without issue in Patagonia¹⁸. This makes their loss, thought to be related to the faunal turnover of the great American exchange beginning approximately 8 Myr ago¹⁹, even more intriguing.

The discovery that the genus *Aotus* has a long independent history corresponds with both the principal serological study of New World monkey phylogeny, which postulated an early divergence from the last common ancestor of the platyrrhine radiation²⁰, and anatomical studies²¹, which predicted an origin for the lineage well before the middle Miocene. No genus of comparable duration is known among the Old World anthropoids, where *Macaca* shows the greatest longevity. It is tentatively traceable to the late Miocene, about 8 Myr ago⁴.

Together with the recent upward revision in the age of the earliest platyrrhine fossil, *Branisella boliviana*, which at 26–27 Myr (ref. 22) is nearly 10 Myr younger than previously supposed, the confirmation of a bathyphyletic modern radiation raises two basic questions. First, has there been a previous adaptive radiation of New World monkeys during the Paleogene? Second, has the modern radiation ever given rise to a significant collateral branch during the Neogene? Thus the implications of dental similarities shared by *Branisella* and modern squirrel monkeys^{1,2,23} need to be reconsidered. Although the fossil's occlusal and mandibular structures retain primitive platyrrhine features²⁴, nothing precludes *Branisella* from occupying a phylogenetic position among the modern cebids². The previously established latest date for the origins of this family is 20 Myr ago²¹, marked by the appearance of *Dolichocebus gaimanensis*, a phyletic link to *Neosaimiri* and *Saimiri*^{3,4}. The available fossil evidence of platyrrhine evolution, therefore, does not support the contention that the rich diversity of the living forms is a function of the group's remote origin. Rather, it suggests a relatively rapid differentiation following a fairly recent origin, and a degree of taxonomic and morphological stasis which far exceeds the pattern found among Old World primates.

We thank INGEOMINAS of the Republic of Colombia for collaborating with us at La Venta and all the members of our field party. Financial support was by a Grant-in-Aid for Overseas Scientific Survey from the Japanese Government and a US–Japan Cooperative Sciences Program grant to the co-authors.

Received 2 December 1986; accepted 6 February 1987.

1. Hershkovitz, P. *Living New World Monkeys (Platyrrhini)* Vol. I (University of Chicago Press, 1977).
2. Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).
3. Rosenberger, A. L. *Nature* **279**, 416–418 (1979).
4. Delson, E. & Rosenberger, A. L. in *Living Fossils* (eds Eldredge, N. & Stanley, S. M.) 50–61 (Springer, New York, 1984).
5. Fields, R. W. *Univ. Calif. Publ. Bull. Dep. Geol.* **32**, 405–444 (1959).
6. Marshall, L. G. in *The Great American Biotic Interchange* (eds Stehli, F. G. & Webb, S. D.) 49–85 (Plenum, New York, 1985).
7. Cartmill, M. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 243–274 (Plenum, New York, 1980).
8. Wright, P. C. in *Ecology and Behavior of Neotropical Primates* (eds Coimbra-Filho, A. F. & Mittermeier, R. A.) 211–240 (Academia Brasileira de Ciencias, Rio de Janeiro, 1981).
9. Kay, R. F. & Cartmill, M. *J. hum. Evol.* **6**, 19–35 (1977).
10. Schultz, A. H. *Am. J. phys. Anthropol.* **26**, 389–408 (1940).
11. Rathbun, G. B. *Primates* **21**, 211–219 (1980).
12. Fleagle, J. G. & Rosenberger, A. L. in *Morphologie Evolutive, Morphogenese du Crâne et Origine de l'Homme* (ed. Sakka, M.) 141–153 (CNRS, Paris, 1983).
13. Fleagle, J. G. & Bown, T. M. *Folia Primatol.* **41**, 249–266 (1983).
14. Rosenberger, A. L. *Phylogeny, Evolution and Classification of New World Monkeys (Platyrrhini, Primates)* (University Microfilms, Ann Arbor, 1979).
15. Marshall, L. G. *J. Paleont.* **50**, 402–418 (1976).
16. Hirschfeld, S. E. *J. Paleont.* **50**, 419–432 (1976).
17. Fields, R. W. *Univ. Calif. Publ. Geol. Sci.* **32**, 273–404 (1957).
18. Cifelli, R. L. in *The Great American Biotic Interchange* (eds Stehli, F. G. & Webb, S. D.) 249–266 (Plenum, New York, 1985).
19. Webb, S. D. in *The Great American Biotic Interchange* (eds Stehli, F. G. & Webb, S. D.) 357–386 (Plenum, New York, 1985).
20. Sarich, V. M. & Cronin, J. R. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 399–421 (Plenum, New York, 1980).
21. Rosenberger, A. L. *J. hum. Evol.* **13**, 737–742 (1984).
22. MacFadden, B. J. *et al. J. Geol.* **93**, 223–250 (1985).
23. Wolff, R. G. *J. vert. Paleont.* **4**, 570–574 (1984).
24. Rosenberger, A. L. *Int. J. Primatol.* **2**, 1–7 (1981).

Early experience of tactile stimulation influences organization of somatic sensory cortex

Daniel J. Simons & Peter W. Land

Department of Physiology and Department of Neurobiology, Anatomy and Cell Science, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, USA

Visual experience is essential for the establishment of the cerebral cortical circuitry that allows normal binocular vision. For example, the pattern of right-eye, left-eye dominance columns is permanently altered by simply closing an eye of a young primate¹. A critical issue is whether environmental factors also influence the development of other cortical sensory areas. In the present experiments we manipulated the tactile experience of young rats by depriving them of the sensory information that is normally provided by their large facial whiskers. Electrophysiological analyses showed that simply trimming the whiskers from the day of birth results in pronounced abnormalities in the response properties of single neurons in the adult somatic sensory cortex. Thus functional plasticity in response to early experience appears to be a fundamental aspect of cortical development.

The somatic sensory cortex of rodents contains discrete groups of neurons in layer IV, called barrels, that are related one-to-one to the large mystacial vibrissae (whiskers) on the contralateral face^{2–4}. Animals had one or more rows of whiskers trimmed for 45–60 days, beginning on the day of birth. In adults this procedure leads to a marked reduction in metabolic activity in the corresponding barrels^{5,6}. Rats were studied electrophysiologically after allowing the whiskers to regrow to normal lengths. Results from a typical experiment are given in Fig. 1. The peristimulus time histograms (PSTs) at the bottom show responses of a neuron recorded extracellularly in the B2 barrel of an animal whose row-C vibrissae had been chronically trimmed. This neuron responded vigorously to deflections of the corresponding vibrissa (hereafter referred to as the Principal Whisker or PW), in this case B2, but poorly to the adjacent, non-trimmed B1 and B3 whiskers. The C2 vibrissa, a chronically trimmed whisker, elicited virtually no response. Also note this neuron's low level of spontaneous activity and, with deflection of B2, its much larger response to stimulus onset than offset. The top PSTs show data obtained from a cell in an homologous part of the adjacent, neonatally deprived C2 barrel of the same animal. Although this neuron, too, responded maximally to stimulation of the barrel's PW (C2), it also responded reliably to deflections of adjacent whiskers, both chronically intact and chronically trimmed. Unlike the neuron in the B2 barrel, this cell responded almost equally well to stimulus onset and offset, and in response to deflection of the C2 whisker the exceptionally high level of spontaneous activity was maintained or slightly elevated during the steady-state phase of whisker deflection. Importantly, these functional changes occurred without any obvious cytoarchitectonic alterations in the barrel field.

Compared to cells in normal barrels, neurons in deprived barrels are more responsive to whisker stimuli, less likely to display strong preferences for whisker movements at a particular angle and respond less selectively to whisker movements in opposing directions. These findings are summarized in Fig. 2, which shows quantitative data derived from PW responses in normal, non-deprived and deprived barrels. The left-hand column shows the magnitudes of ON responses elicited by whisker movements at the angle that was most effective in activating an individual neuron. Histograms in the remaining columns indicate how selective the cells were to whisker movements at different angles. Statistical comparisons (*t*-tests or χ^2

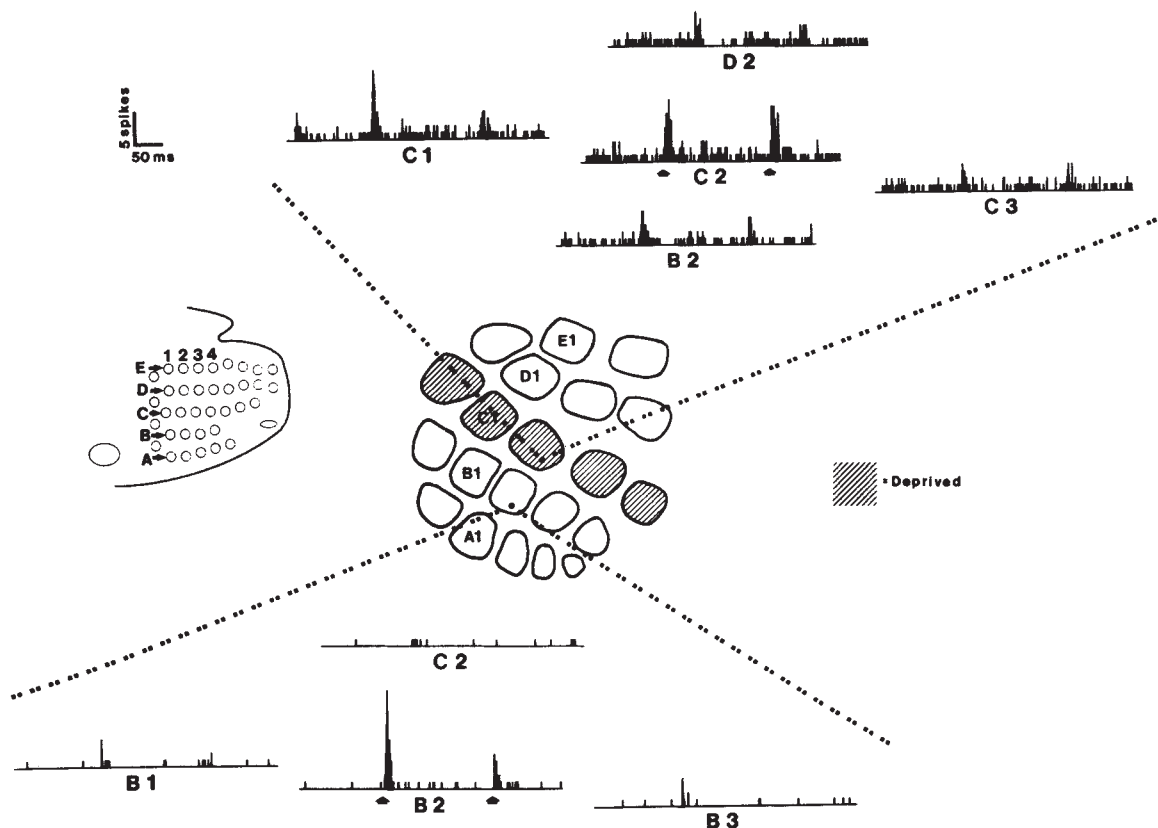


Fig. 1 Comparison of responses elicited by deflection of the Principal Whisker (PW) and adjacent whiskers for a neuron recorded in a deprived barrel and one recorded in an homologous part of a non-deprived barrel in the same animal. The left row-C whiskers had been trimmed from birth until 54 days of age, and the recordings were obtained 83 days later. The centre shows a reconstruction of a tangential section through layer IV of this animal's right somatic sensory cortex. The cytochrome oxidase-rich centres of barrels visible in this section are outlined and the adjacent diagram illustrates the arrangement of the whiskers on the left face. Peristimulus time histograms (PSTs) at the top show responses of a neuron in the deprived C2 barrel to mechanically controlled deflections of its regrown PW (C2) and of 4 adjacent whiskers. The individual vibrissae were deflected randomly in eight angular directions (in 45° increments relative to the horizontal alignment of the whisker rows), and for this cell each series was repeated 5 times for a total of 40 stimulus presentations. Stimuli consisted of ramp-and-hold 'trapezoids' of 200 ms duration; stimulus onset and offset are indicated by arrows for the PW response. PSTs at the bottom show responses of a neuron in the non-deprived B2 barrel to deflection of its PW (B2) and 3 adjacent vibrissae.

Methods. Data were obtained from adult male and female Long-Evans rats. Under ether anaesthesia rats had either row C whiskers ($n=7$) or all whiskers except those in row C ($n=2$) on the left face trimmed to within 1 mm of the skin surface from birth to 45–60 days of age, after which they were allowed to regrow for 3–15 weeks. Within 4 weeks, even the largest hairs had grown to 75% of their expected adult lengths. Deprived barrels, barrels corresponding to chronically trimmed whiskers; non-deprived barrels, adjacent barrels associated with chronically intact whiskers. Five untrimmed rats served as normal controls. Rats were anaesthetized with sodium pentobarbital or methohexital and prepared for electrophysiological study as described⁸. During the recording sessions the rats were maintained under fentanyl, an opiate analogue (Sublimaze, 5–10 µg per kg per h), administered pancuronium bromide to prevent movement or opiate-induced piloerection of the vibrissae, and artificially respired. The condition of the rats was assessed by observation of the electroencephalogram, electrocardiogram, expired carbon dioxide levels, pupillary reflex, and capillary perfusion of glabrous skin. Double-barrelled glass micropipettes were used to examine the extracellularly recorded activity of individual neurons in the right hemisphere. Electrode tracks were marked by iontophoretically ejecting horseradish peroxidase (HRP) from one side of the micropipette. Electromechanical stimulators produced controlled deflections of individual vibrissae on the left face²². After the recording sessions rats were killed and their brains processed for cytochrome oxidase and HRP histochemistry. All electrode tracks were reconstructed histologically using serial tangential sections through the barrel field. Data are reported only from those units recorded in the cytochrome oxidase-rich centres of the barrels²³.

tests) showed that the population of neurons in deprived barrels differs from that of neurons in normal barrels on all three measures ($P \leq 0.05$). Cells in non-deprived barrels are not entirely normal either. On measures of maximal discharge and angular tuning they display intermediate values that are not significantly different from normal ($P > 0.05$); their ON/OFF ratios are, however, significantly smaller than those of neurons in normal barrels ($P \leq 0.05$) and similar to those of deprived barrel neurons.

Figure 3 shows population profiles representing the summed responses of all cells from each experimental group. ON and OFF responses are clearly largest in deprived barrels. Bin-by-bin statistical analyses revealed that differences between deprived and normal barrel neurons are present even in the first 2 ms of the ON response (t -test, $P \leq 0.05$). Deprived neurons also dis-

play more spontaneous discharges and longer duration ON and OFF responses. Moreover, during the period when the whisker is deflected subsequent excitatory peaks are more frequent and activity does not fall below spontaneous levels. By contrast, activity in normal barrels is reduced immediately after the ON response, and over a period of 75 ms it gradually returns to, then transiently exceeds, pre-stimulus levels. This reduction of activity parallels the time course of intracellularly recorded inhibitory postsynaptic potentials⁷ and coincides with the period during which neuronal discharges to movements of a second, adjacent whisker can be suppressed⁸. Interestingly, the first peak following the ON response of deprived neurons occurs during the period when activity in normal cells is most strongly inhibited. In non-deprived barrels, too, the first major peak following the ON response occurs earlier than normal. These

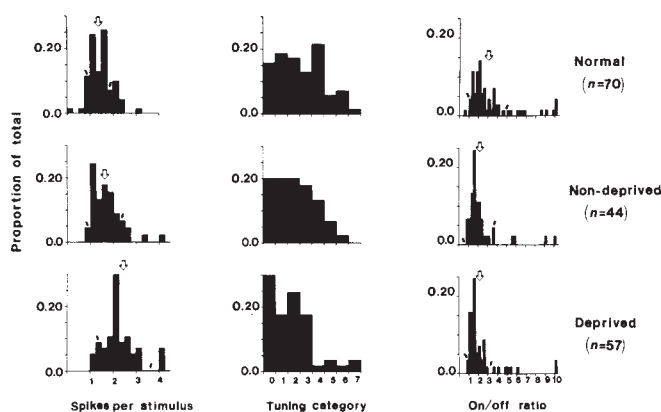


Fig. 2 Response properties of barrel neurons recorded in normal animals and in non-deprived and deprived barrels of experimental animals. Data are derived from responses elicited by stimulating the whisker that corresponded anatomically to the barrel in which a neuron was located (PW responses); n , number of neurons in each group. Left, size of ON responses elicited by whisker movements at each cell's maximally activating or best angle. ON responses are computed as the average number of spikes per stimulus elicited during the 20 ms following the onset of whisker movement. Right-most bin on abscissa shows all values ≥ 4.0 . Open arrowhead, mean; solid arrows, standard deviation. Middle, angular tuning histograms. Each cell is classified on the basis of how many angles elicited ON responses statistically smaller than those at the maximally activating angle (t -tests, $P \leq 0.05$). Category 0 represents non-tuned cells, category 7 the most highly tuned cells. To compare these data from different experimental groups χ^2 tests were used (see text). Right, ON/OFF ratios. PSTs such as those shown in Fig. 1 were used to compare the relative magnitudes of ON and OFF responses. The number of spikes evoked by moving the whisker to its deflected state (ON responses) was divided by the number of spikes evoked by returning the whisker to its resting or neutral position (OFF responses). Right-most bin on abscissa shows all ratios ≥ 10.0 . Because of some extremely large values in the normal group, means (open arrowheads) and standard deviations (solid arrows) were computed exclusive of ratios ≥ 10.0 .

findings suggest that cortical neurons in deprived animals probably display abnormal responses to temporally patterned stimuli involving sequential deflections of adjacent vibrissae.

Simple whisker trimming in young rats leads to enlarged receptive fields, reduced angular tuning, increased responsiveness and altered temporal patterns of stimulus-evoked discharges of cortical neurons. Such abnormalities were observed in all trimmed animals indicating that up to 3 months of experience with regrown whiskers was not sufficient to reverse the effects of the neonatal deprivation. In spite of these alterations, basic vibrissa/barrel relations are maintained in the deprived rats⁹. These findings contrast with those from visual system studies in which deprived neurons become unresponsive and/or dominated by inputs from the non-deprived eye¹⁰⁻¹². It may be significant that the cytoarchitectonic differentiation of barrels occurs independently of functional activity.

One explanation that could account for the observed functional changes is a general increase in neuronal excitability, a possibility consistent with the observed elevation of spontaneous activity in deprived barrels. Although cell-by-cell analyses of deprived barrels showed that spontaneous activity is in fact highly and positively correlated with elevated activity during the maintained phase of whisker deflection ($r = 0.77$, $P \leq 0.001$), it is not significantly correlated with activity during the first 2 ms of the ON response ($r = 0.20$, $P > 0.05$). Further, cells that were more strongly driven by direct thalamic input, as measured by their larger initial responses during the first 2 ms, did not necessarily continue to respond vigorously during sustained whisker displacements ($r = -0.01$). These analyses argue against a general increase in excitability and suggest that two separate,

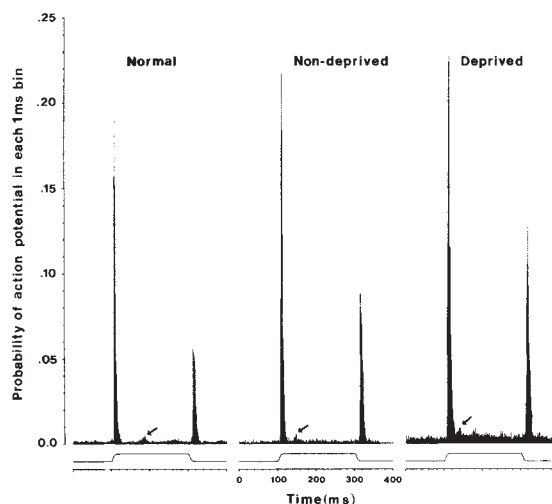


Fig. 3 Population profiles showing the summed responses of all neurons in each experimental group. Histograms were constructed by a bin-by-bin accumulation of responses to all deflections of the PWs. Total number of stimulus presentations: normal, 5,040; non-deprived, 2,640; and deprived, 2,280. Data are expressed as the observed probability that an action potential occurred within that 1 ms bin. Stimulus waveforms and timescales are shown below each PST and the arrowhead denotes the first major excitatory peak following the ON response.

though not necessarily independent, mechanisms may be responsible for the overall increase in the responsiveness of deprived barrel neurons. We propose that the present findings reflect increased or more efficient input from the thalamus to cortical neurons together with diminished effectiveness of the intracortical inhibitory mechanisms that normally depress spontaneous activity of cortical neurons and shape their responses to maintained whisker displacements.

Previous studies of the somatic sensory cortex have shown that topographical maps of the contralateral body surface are altered by nerve section or digit removal¹³. In rodents, direct damage to the whisker nerves early in life produces parallel alterations in barrel cytoarchitecture and cortical somatotopy^{14,15}. All these manipulations involve peripheral deafferentation. The finding that a relatively innocuous and non-invasive procedure such as whisker trimming can markedly alter the dynamic response properties of cortical neurons indicates that functional activity *per se* is critically important in the development of somatic sensory cortex. We have not yet determined whether sensory deprivation in adult animals results in similar electrophysiological changes, but our conclusions are consistent with studies showing increased glucose use in the somatic sensory cortex following chronic whisker trimming in neonatal but not adult rats⁹.

The rodent cortex remains immature for up to six weeks after birth¹⁶⁻²⁰, and normal 'whisking' behaviour appears during the third postnatal week²¹. Thus the intrinsic organization of the barrel field is still developing when it first receives adult-like patterns of tactile input. The pronounced functional abnormalities in cortical neurons produced by neonatal sensory deprivation may reflect an exaggerated outcome of a process that, under normal circumstances, enables the brain to adapt to an external environment that must be perceived in new and subtly different ways as the young animal becomes more mobile and adept at interacting with its surroundings.

This work was supported by the NIH and the Pittsburgh Chapter of the United Way. We thank Glenn Fleet for expert technical assistance, Ellen McCracken for typing the manuscript and Drs George Carvell, Eric Frank and John Horn for helpful comments and suggestions.

Received 8 January; accepted 5 February 1987.

1. Hubel, D. H., Wiesel, T. N. & LeVay, S. *Phil. Trans. R. Soc. B* **278**, 377-409 (1977).
2. Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17**, 205-242 (1970).
3. Welker, C. *Brain Res.* **26**, 259-275 (1971).
4. Welker, C. *J. comp. Neurol.* **166**, 173-190 (1976).
5. Land, P. W. & Simons, D. J. *Brain Res.* **341**, 189-194 (1985).
6. Durham, D. & Woolsey, T. A. *J. comp. Neurol.* **178**, 629-644 (1978).
7. Carvell, G. E. & Simons, D. J. *Soc. Neurosci. Abstr.* **11**, 751 (1985).
8. Simons, D. J. *J. Neurophysiol.* **54**, 615-635 (1985).
9. Hand, P. J. in *Changing Concepts of the Nervous System* (eds Morrison, A. R. & Strick, P. L.) 49-68 (Academic, New York, 1982).
10. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **28**, 1029-1046 (1965).
11. LeVay, S., Wiesel, T. N. & Hubel, D. H. *J. comp. Neurol.* **191**, 1-51 (1980).
12. Dräger, U. C. *J. Neurophysiol.* **41**, 28-42 (1978).
13. Kaas, J. H., Merzenich, M. M. & Kilackey, H. P. A. *Rev. Neurosci.* **6**, 325-356 (1983).
14. Durham, D. & Woolsey, T. A. *J. comp. Neurol.* **235**, 97-110 (1985).
15. Simons, D. J., Durham, D. & Woolsey, T. A. *Somatosens. Res.* **1**, 207-245 (1984).
16. Rice, F. L. *J. comp. Neurol.* **236**, 496-503 (1985).
17. Waite, P. M. E. *J. comp. Neurol.* **173**, 165-174 (1977).
18. Armstrong-James, M. *J. Physiol., Lond.* **246**, 501-538 (1975).
19. Coyle, J. T. & Enna, S. J. *Brain Res.* **111**, 119-133 (1976).
20. Kriegstein, A. R., Suppes, T. & Prince, D. *Soc. Neurosci. Abstr.* **9**, 602 (1983).
21. Welker, W. I. *Behaviour* **22**, 223-244 (1964).
22. Simons, D. J. *Brain Res.* **276**, 178-182 (1983).
23. Land, P. W. & Simons, D. J. *J. comp. Neurol.* **238**, 225-235 (1985).

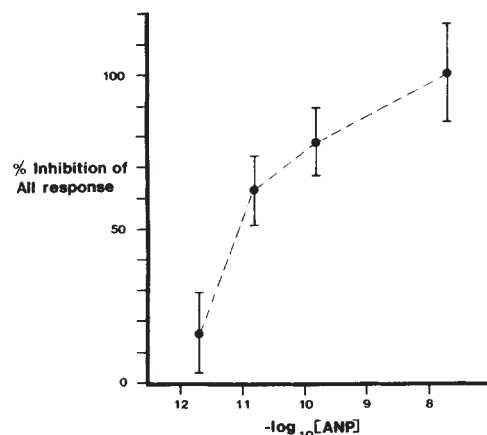


Fig. 1 Dose-response of inhibition of angiotensin-stimulated proximal fluid reabsorption by addition of α -rANP₁₋₂₃ to peritubular perfusion fluid containing AII (1.1×10^{-12} M). Values are means $\pm$ s.e.m.

Atrial natriuretic peptide inhibits angiotensin-stimulated proximal tubular sodium and water reabsorption

Peter J. Harris, David Thomas & Trefor O. Morgan

Department of Physiology, University of Melbourne, Parkville, Victoria 3052, Australia

The discovery that atrial extracts have potent diuretic and natriuretic properties¹ revealed a possible endocrine function of the heart in the regulation of extracellular fluid volume. Since that first report intense research activity has been directed towards determining the mechanism of action of the active atrial natriuretic polypeptides (ANP) found in these extracts. Despite these efforts it remains controversial whether the renal actions of ANP are exerted solely on the process of glomerular filtration or involve additional direct actions on tubular transport. We have investigated the possibility that atrial natriuretic polypeptides may induce natriuresis by suppression of proximal tubular sodium and water reabsorption. Using shrinking split-drop micropuncture combined with simultaneous capillary perfusion in anaesthetized rats we report that 20 nanomolar α -rANP₁₋₂₃ (the main component of ANP in rat plasma) added to the peritubular fluid had no direct effect on proximal fluid uptake whereas picomolar angiotensin II had a marked stimulatory action as reported². The stimulatory effect of angiotensin II on fluid reabsorption was inhibited by peritubular ANP at physiological concentrations and abolished by higher concentrations of ANP. Thus at physiological concentrations^{3,4} ANP acts within the kidney to decrease proximal reabsorption by inhibition of angiotensin-stimulated sodium and water transport.

Most studies indicate that at high concentrations ANP raises glomerular filtration rate. However, although this mechanism has been considered by some to provide an adequate explanation for the natriuretic action of ANP⁵, others have found evidence for both haemodynamic and tubular transport actions^{6,7}. A direct action on tubular transport is supported by the finding that lower doses increase sodium excretion without detectable changes in glomerular filtration rate (GFR)⁸ but more direct studies on isolated perfused segments of rabbit proximal tubule⁹ and loop of Henle¹⁰ failed to demonstrate a tubular site of action. A possible answer lies in reports of interactions between ANP and angiotensin II (AII) in a variety of tissues. ANP inhibits the constrictor actions of AII on blood vessels *in vitro*¹¹, the systemic pressor action¹² and AII-stimulated aldosterone synthesis¹³. We reasoned that ANP may inhibit the stimulation

of proximal tubular sodium and water transport induced by peritubular AII at physiological concentrations^{2,14}.

In butabarbital-anaesthetized male Sprague-Dawley rats prepared for micropuncture of the left kidney¹⁵ and infused with 0.9% NaCl at 4.2 ml h^{-1} , proximal tubular fluid reabsorption rates were measured by a computerized imaging modification of the shrinking split-drop micropuncture technique¹⁶. A droplet of castor oil stained with Sudan Black was placed in the tubule through one barrel of a double-barrelled micropipette and from the other barrel artificial tubular fluid was introduced to split the oil column. The rate of shrinking of the split-drop was determined by digital computer analysis of the positions of the oil-water menisci in successive video frames captured at one-second intervals. Adjacent capillaries were simultaneously perfused with a fluid of plasma-like composition². Perfusion started at least 10 s before introducing the tubular fluid and was driven by gas (95% O₂/5% CO₂) applied to the micropipette at a pressure sufficient to clear the blood from all capillaries adjacent to the split-drop.

The results are shown in Table 1. Addition of 2×10^{-8} M α -rANP₁₋₂₃ (this peptide, atriopeptin II, comprises the residues Ser128-Arg150 of the preatriopeptin sequence and is the main component of ANP in rat plasma) to the peritubular perfusate had no effect on fluid reabsorption compared with control (group 1). In groups 2-5, peritubular AII (1.1×10^{-12} M) decreased shrinking drop half-time ($t_{1/2}$) indicating stimulation of proximal tubular fluid reabsorption rate (expressed per unit area of epithelium). When ANP (1.8×10^{-12} M- 1.8×10^{-8} M) was added to the peritubular perfusate together with AII (1.1×10^{-12} M), reabsorption decreased in a dose-dependent manner (Fig. 1). At a concentration of 1.8×10^{-8} M values of reabsorption were similar to those observed with perfusate alone, indicating complete abolition of the stimulatory effect of AII. These data confirm previous observations that ANP has no direct action by itself on proximal tubular transport^{9,16} but also demonstrate the ability of physiological doses of ANP to suppress proximal fluid reabsorption by inhibition of AII-stimulated sodium transport. The mechanisms involved in this action are unknown.

The proximal tubular action of ANP reported here provides a ready explanation for the increase in fractional lithium and phosphate excretion found in anaesthetized dogs⁶ and taken to indicate inhibition of proximal sodium and water reabsorption. Cogan⁵ also observed reduced fractional proximal reabsorption in anaesthetized rats as absolute proximal reabsorption did not rise in parallel with GFR. He noted an increase of 30-35% in absolute delivery of sodium, bicarbonate and chloride ions out of the proximal tubule and commented that this was responsible, at least in part, for the observed natriuresis. However, sub-

Table 1 The effects of AII (1.1×10^{-12} M) and α -rANP₁₋₂₃ proximal tubular fluid reabsorption

Peritubular perfusion fluid	Proximal tubular fluid reabsorption ($\times 10^{-4}$ mm ³ mm ⁻² s ⁻¹)
Group 1 (n = 5)	
Control	2.43 $\pm$ 0.20
ANP (2×10^{-8} M)	2.61 $\pm$ 0.27
Group 2 (n = 5)	
Control	1.99 $\pm$ 0.17
AII	3.11 $\pm$ 0.28*
AII + ANP (1.8×10^{-12} M)	2.82 $\pm$ 0.94†
Group 3 (n = 5)	
Control	2.20 $\pm$ 0.15
AII	2.93 $\pm$ 0.22*
AII + ANP (1.8×10^{-11} M)	2.43 $\pm$ 0.10†
Group 4 (n = 5)	
Control	2.02 $\pm$ 0.21
AII	3.13 $\pm$ 0.39*
AII + ANP (1.8×10^{-10} M)	2.30 $\pm$ 0.18
Group 5 (n = 6)	
Control	2.35 $\pm$ 0.06
AII	3.42 $\pm$ 0.32*
AII + ANP (1.8×10^{-8} M)	2.35 $\pm$ 0.22

* Significantly different ($P < 0.05$; paired *t*-test) from control.

† Significantly different from AII value.

Each value represents the mean $\pm$ s.e.m. of the average of measurements from 3–5 randomly punctured tubules with each solution in each animal; *n*, number of animals.

sequent return of 90% effective proximal glomerulotubular balance and abolition of the natriuretic response during reduction of GFR by aortic constriction led him to conclude that ANP acts in the kidney predominantly by raising GFR. The remaining discrepancy between solute excretion and filtered load could not be attributed to redistribution of renal blood flow and no tubular site of action had been positively demonstrated. Inhibition of angiotensin-stimulated fluid reabsorption can account fully for the results of these experiments. In addition, because the magnitude of the natriuresis depends on the change in filtered load, the peritubular AII concentration and the extent of compensatory distal reabsorption, the considerable variability in the natriuretic response to ANP¹⁷ could reflect differing AII levels in various states of sodium balance.

Inhibition of the actions of AII by ANP may also contribute to natriuresis by two additional complementary mechanisms in the kidney. ANP inhibits renin release from isolated rat juxtaglomerular cells¹⁸ and would therefore be expected to reduce peritubular AII levels with a consequent reduction in angiotensin-stimulated transport¹⁷. Second, blunting of the tubuloglomerular feedback response¹⁹ would facilitate the natriuretic response to ANP. Maintenance of GFR, despite a rise in distal solute delivery caused by increased filtered load combined with decreased fractional proximal reabsorption, would enhance urine flow and solute excretion.

ANP was supplied by Dr F. Spinelli of Ciba-Geigy, Basle. The study was supported by the National Health and Medical Research Council of Australia.

Received 8 December 1986; accepted 4 February 1987.

- DeBold, A. J., Borenstein, H. B., Veress, A. T. & Sonnenberg, H. *Life Sci.* **28**, 89–94 (1981).
- Harris, P. J. & Young, J. A. *Pflügers Arch. ges. Physiol.* **367**, 295–297 (1977).
- Tanaka, I., Misono, K. S. & Ingami, T. *Biochem. biophys. Res. Commun.* **124**, 663–668 (1984).
- Miyata, A., Kangawa, K., Toshimori, T., Hatoh, T. & Matsuo, H. *Biochem. biophys. Res. Commun.* **129**, 248–255 (1985).
- Cogan, M. G. *Am. J. Physiol.* **250**, F710–F714 (1986).
- Burnett, J. C. Jr, Granger, J. P. & Oppenorth, T. J. *Am. J. Physiol.* **247**, F863–F866 (1984).
- Burnett, J. C. Jr, Oppenorth, T. J. & Granger, J. P. *Kidney Int.* **30**, 16–19 (1986).
- Gellai, M., DeWolf, R. E., Kinter, L. B. & Beuwer, R. III *Circulation Res.* **59**, 56–62 (1986).

- Baum, M. & Toto, R. D. *Am. J. Physiol.* **250**, F66–F69 (1986).
- Kondo, Y., Imai, M., Kangawa, K. & Matsuo, H. *Pflügers Arch. ges. Physiol.* **406**, 273–278 (1986).
- Kleinert, H. D. *et al. Hypertension* **6** (suppl. 1), I143–I147 (1984).
- Di Nicolantonio, R., Stevens, J., Weaver, D. & Morgan, T. O. *Clin. exp. Pharmac. Physiol.* **13**, 311–315 (1986).
- Atarashi, K., Mulrow, P. J., Franco-Saenz, R., Snajdar, R. & Rapp, J. *Science* **224**, 992–994 (1984).
- Schuster, V. L., Kokko, J. P. & Jacobson, H. R. *J. clin. Invest.* **73**, 507–515 (1984).
- Gyory, A. Z., Lingard, J. M., & Young, J. A. *Pflügers Arch. ges. Physiol.* **348**, 205–210 (1974).
- Harris, P. J., Cullinan, M. & Thomas, D. *Pflügers Arch. ges. Physiol.* (in the press).
- Spinelli, F., Kamber, B. & Schnell, C. *J. Hypertension* **4** (suppl. 2), S25–S29 (1986).
- Kurtz, A., Bruna, R. D., Pfeilschifter, J. & Bauer, C. *J. Hypertension* **4** (suppl. 2) S57–S60 (1986).
- Huang, C.-L., Ives, H. E. & Cogan, M. G. *Kidney Int.* **29**, 384 (1986).

Activation of NMDA receptors blocks GABAergic inhibition in an *in vitro* model of epilepsy

Armin Stelzer*, N. Traverse Slater* & Gerrit ten Bruggencate

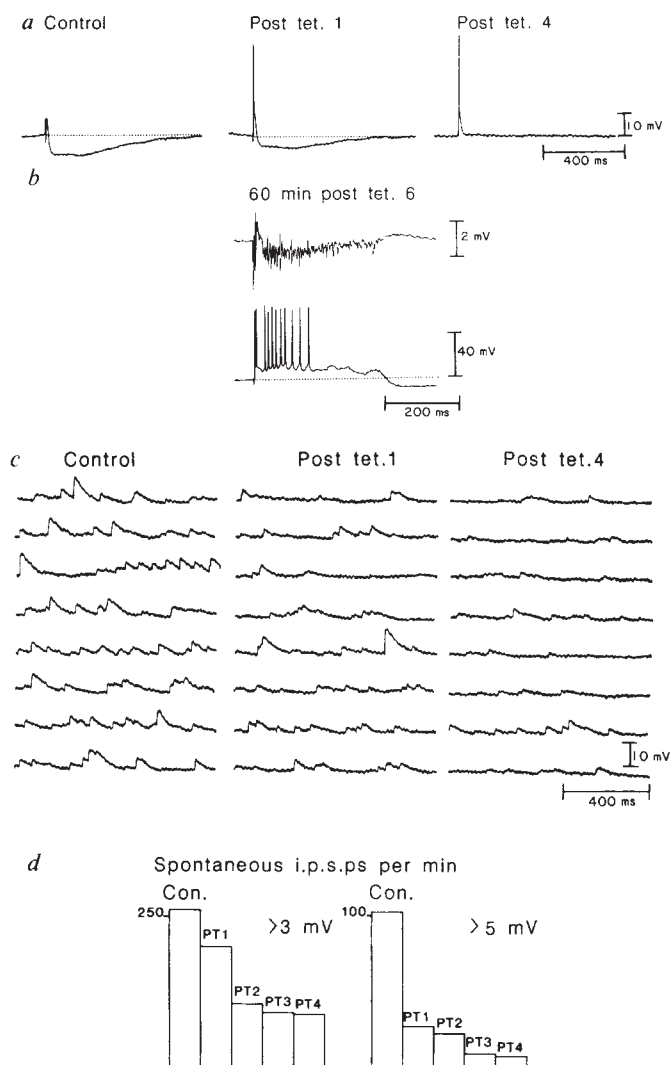
Physiology Institute, University of Munich, Pettenkoferstrasse 12, D-8000 Munich 2, FRG

The application of tetanic electrical stimuli to the stratum radiatum fibre pathway in the hippocampus *in vitro* produces an NMDA (*N*-methyl-D-aspartate) receptor-dependent enhancement of synaptic efficacy^{1–3}. Repeated application of such stimuli produces a progressive enhancement of synaptic efficacy leading to the genesis of spontaneous and stimulation-evoked epileptiform discharges^{4,5}. We have used this *in vitro* approach to explore the cellular mechanisms which underlie the kindling model of epilepsy. Kindling of the stratum radiatum fibre pathway *in vitro* induced a progressive, long-lasting reduction of both spontaneous and stimulation-evoked GABAergic (γ -aminobutyric acid-mediated) inhibitory postsynaptic potentials (i.p.s.ps). The reduction of i.p.s.ps by kindling was associated with a profound decrease in the sensitivity of CA1 pyramidal neurons to ionophoretically applied GABA and an increase in sensitivity to NMDA. The reduction of i.p.s.ps and GABA sensitivity was prevented by kindling in the presence of the NMDA receptor antagonist D-2-amino-5-phosphonovalerate (D-APV). These results demonstrate that kindling-like stimulus patterns produce a reduction of GABAergic inhibition in the hippocampus resulting from a stimulus-induced postsynaptic activation of NMDA receptors. The modulation of GABAergic inhibition by NMDA receptors may cause the synaptic plasticity which underlies the kindling model of epilepsy.

Experiments were performed on transverse slices (0.5 mm thick) of guinea-pig hippocampus. Slices were cut with a vibrating tissue cutter and maintained submerged in an experimental chamber perfused with a solution of the following composition (mM): NaCl 118, KCl 3, NaHCO₃ 25, NaH₂PO₄ 1.2, MgCl₂ 1, CaCl₂ 1.5, D-glucose 11. The bath temperature was kept at 29–33 °C, and the solution constantly gassed with 95% O₂/5% CO₂. Intra- and extracellular recordings were made in the CA1 stratum pyramidale of slices in which the CA2/3 subfield was removed by dissection⁴. Evoked responses were produced by stimulation of stratum radiatum Schaffer collateral/commissural fibres through a pair of sharpened, insulated tungsten electrodes. Evoked intracellular potentials were recorded either with microelectrodes filled with 4 M potassium acetate (40–100 MΩ resistance), or with microelectrodes filled with 3 M potassium chloride (30–60 MΩ resistance; to measure spontaneous inhibitory post-

* Present addresses: Department of Neurology, Room 4-408, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA (A.S., to whom correspondence should be addressed); Departments of Physiology and Psychiatry, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, Illinois 60611, USA (N.T.S.).

Fig. 1 Effects of repeated tetani on orthodromically evoked post-synaptic potentials (*a, b*) and on spontaneous depolarizing i.p.s.ps (*c, d*). *a*, Progressive decline of both phases of orthodromically evoked i.p.s.ps by repetitive high-frequency stimulation. *b*, Paroxysmal depolarization shift recorded extra- (upper trace) and intracellularly (lower trace) triggered by the epileptiform orthodromic e.p.s.p. 60 min after the sixth tetanus had been applied. Amplitudes of action potentials in *a* and *b* are attenuated. *c, d*, Tetanization-induced progressive reduction of spontaneous depolarizing i.p.s.ps recorded in another cell impaled with a 3 M KCl-filled micropipette. Each column of traces in *c*, contains representative sweeps recorded at the three points in time: control, after tetanus 1 (post tet. 1) and after tetanus 4 (post tet. 4). The histograms in *d* summarize the results from the cell shown in *c*, PT, post tet.; Con, control. The number of depolarizing events greater than 3 mV (left) or 5 mV (right) in amplitude recorded over a 1-min period beginning 20 min after each tetanus is shown. Results similar to those in *c* and *d* were observed in five other CA1 pyramidal cells. Evidence that the depolarizing events were GABA-mediated was obtained by their reversible blockade following application of bicuculline-methiodide (10^{-4} – 10^{-5} M) (not shown). The current-voltage relationships for both cells illustrated remained unchanged during the experiment. Resting membrane potentials: *a*, -70 mV; *c*, -65 mV; input resistance: *a*, 35 M Ω ; *c*, 80 M Ω .



synaptic potentials⁶). Drugs (D-APV, bicuculline-methiodide) were applied in fixed concentrations by bath perfusion. GABA (1 M, pH 3.5) and NMDA (50 mM, pH 7.0) were applied through an extracellular ionophoretic electrode (three-barrelled, third channel filled with 0.9% NaCl for current balancing) to the edge of stratum pyramidale. The position of the ionophoretic pipette and the intensity of the GABA ejecting current were adjusted to elicit biphasic GABA responses^{7,8}, whenever possible at resting membrane potentials. GABA was ejected with positive current pulses (10–155 nA, retaining current 3–10 nA), NMDA with negative current pulses (20–120 nA, retaining current 3–10 nA). Data were digitized and stored on disk for subsequent off-line analysis.

Tetanic orthodromic stimulation consisted of 5 trains (50 Hz, 2 s) delivered at 10 s intervals⁴. The stimulus strength was set at twice the intensity which produced a maximal orthodromic extracellular population spike in each slice (range 4–10 V, single pulse duration 100 μ s). This stimulation procedure was repeated every 30 min, and data were collected 20 min after each preceding tetanus.

Figure 1*a* shows that the application of tetanic stimulation every 30 min produced a progressive decrease in the magnitude of both phases of the orthodromically evoked i.p.s.ps. After four tetani, orthodromically evoked i.p.s.ps were always reduced to 25% or less of control values. The voltage changes induced by hyperpolarizing and depolarizing intracellular current pulses were not affected by high-frequency stimulation (Fig. 3*a*, see ref. 4), indicating that this marked decrease of i.p.s.p. was not

associated with alterations of membrane properties such as resting potential, input resistance, inward rectification, spike accommodation or afterhyperpolarization. Spontaneous depolarizing i.p.s.ps recorded with 3 M KCl-filled electrodes were similarly affected. The amplitude and apparent frequency was progressively reduced to about 25% or less of control values after four tetani (Fig. 1*c, d*). The reduction of orthodromically evoked or spontaneous i.p.s.ps was observed in all (more than 20) cells studied under these conditions, although there was some variability in the control size of i.p.s.ps and in the development of i.p.s.p. reduction following tetanic stimulation.

The enhanced synaptic excitability after the first tetanus was matched by a reduction of both evoked and spontaneous i.p.s.ps of about 50% (Fig. 1*a, d*; and Fig. 3*b*). Epileptiform activity, including paroxysmal depolarization shifts (PDSs), occurred after i.p.s.ps had been reduced to less than 10% of control values. This usually required four or more tetani in cells with well developed control i.p.s.ps (>6 mV). However, in neurons exhibiting only small or no measurable i.p.s.ps under control conditions, epileptiform discharges appeared often after the second tetanus had been applied. The enhanced excitation and decreased inhibition induced by tetanization was always evident until the end of the experiment (up to four hours following the application of the first tetanus). The measurement of the e.p.s.p.-i.p.s.p. sequence at various membrane potentials (by depolarizing or hyperpolarizing d.c.-current application) revealed a proportional reduction of i.p.s.p. amplitudes at each membrane potential following repeated high-frequency stimulation; the

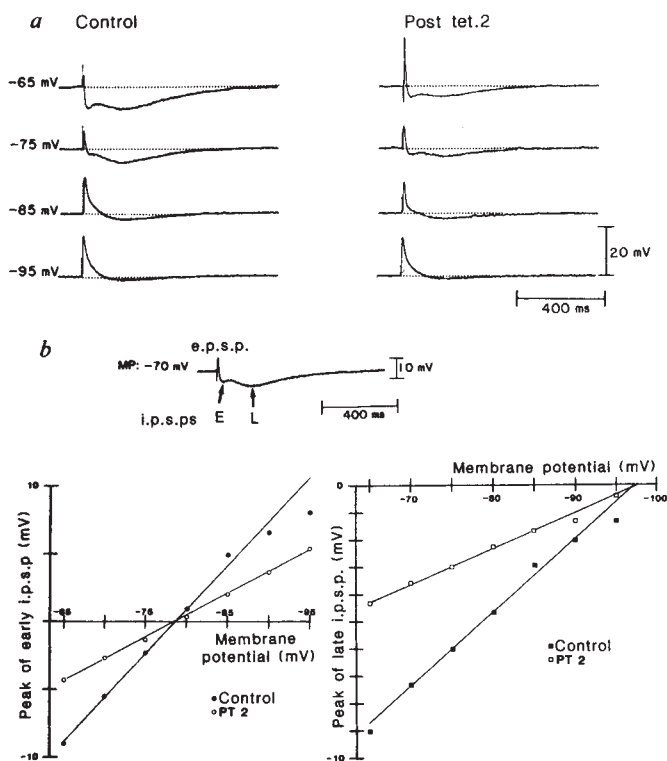


Fig. 2 *a*, Orthodromically evoked e.p.s.p.-i.p.s.p. sequence at various membrane potentials before tetanization and after the second tetanus had been applied. *b*, Trace (above) illustrating early phase of i.p.s.ps mediated by Cl^- currents²⁰ (arrow E, below trace) and late phase of i.p.s.ps mediated by K^+ currents²¹ (arrow L, below trace). Latencies to peak: early, 30 ms; late, 180 ms. MP, membrane potential. Below, graphs of peak of i.p.s.p. and membrane potential for early (left-hand graph, circles) and late (right-hand graph, squares) i.p.s.ps: filled symbols, controls; open symbols, after tetanus 2. Kindling-like stimuli produced a proportional reduction of both phases at each potential recorded (from -65 to -95 mV). The reversal potentials of both i.p.s.p. components were unaltered following tetanic stimulation in all cells examined (Control: early, -74.9 ± 3.0 ($n=8$); late, -97.2 ± 4.6 ($n=6$), after first tetanus: early, -75.5 ± 3.1 ($n=8$); late, -94.4 ± 4.6 ($n=5$), after second tetanus: early, -73.9 ± 4.1 ($n=6$); late, -98.0 ± 4.5 ($n=4$); means $\pm$ s.e.m.). Reversal potentials of each phase were estimated from straight lines drawn by eye. Resting membrane potential, -66 mV; input resistance, 35 M Ω .

reversal potential of both phases of the orthodromically evoked i.p.s.p. was not significantly altered ($P > 0.1$, paired t -test, Fig. 2). This lack of change in the reversal potential of either phase of the orthodromic i.p.s.p. indicates that the tetanus-induced reduction of i.p.s.ps observed here was not due to the changes in ionic equilibria which probably underlie the transient reduction of i.p.s.ps in these cells immediately following tetani^{7,9,10}.

In the presence of the NMDA receptor antagonist D-APV (5–10 μM), tetanic stimuli did not affect excitatory or inhibitory postsynaptic potentials (Fig. 3). Following the washout of D-APV, further tetani produced a reduction of i.p.s.p. amplitude and an increase of e.p.s.p. amplitude. The prolonged epileptiform e.p.s.p. in kindled slices displays a late (> 10 ms), NMDA-receptor-mediated component which can be reversibly antagonized by D-APV in this dose range⁴. A similar contribution of a late, slow NMDA-receptor-mediated component to the e.p.s.p. evoked by Schaffer collateral/commissural fibre stimulation has been demonstrated in the presence of convulsant drugs^{11,12}. This dose range of D-APV was also quite selective in protecting i.p.s.ps from the effects of tetani, without direct actions on

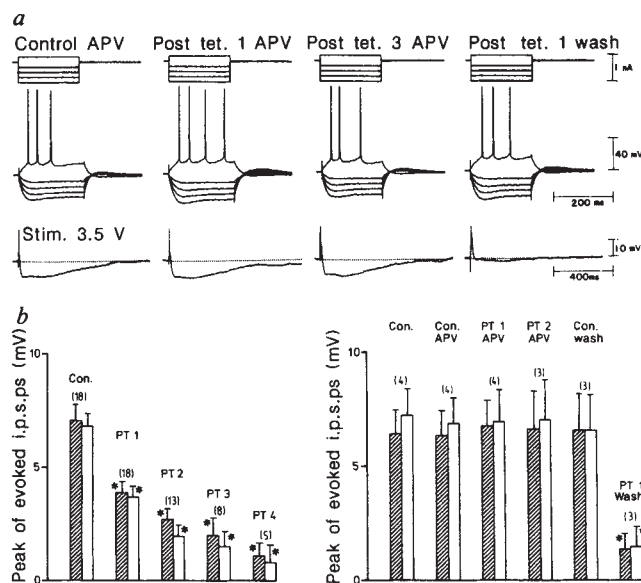


Fig. 3 Effects of D-APV (1×10^{-5} M) on orthodromically evoked synaptic potentials following tetanization. *a*, Evoked responses to stratum radiatum stimulation (3.5 V) were recorded intracellularly in CA1 stratum pyramidale in D-APV-containing control solution after the first and after the third tetanus (post tet. 1 and post tet. 3), and after the first tetanus following the washout of D-APV (wash). *b*, The left histogram is of i.p.s.p. peak sizes (both early (shaded bars) and late (open bars) phases, see inset in Fig. 2*b*, from control until 30 min after the fourth tetanus had been applied) in cells of slices which were not exposed to D-APV; the right histogram is the same in slices exposed to D-APV during tetanization, after washout and further tetanization (vertical bars, s.e.m.; *, $P < 0.001$ in a paired t -test; numbers in parentheses, number of cells). At this concentration D-APV did not have any effect on inhibitory synaptic potentials, but did reversibly suppress the otherwise observed changes in the evoked e.p.s.p.-i.p.s.p. response following tetanic stimulation. Neither tetanization itself, nor the application of D-APV at this concentration, nor the combination of APV and tetanic stimulation had any effect on the current-voltage relationship obtained by intracellular injection of short hyper- and depolarizing current pulses (*a*). The reduction of orthodromically evoked i.p.s.ps after washout of D-APV and further tetanization was somewhat faster than that in cells which were not previously tetanized in D-APV (*a* and *b*). Amplitudes of action potentials in *a* are attenuated. Resting membrane potential of the cell shown in *a*, -67 mV; input resistance, 45 M Ω .

pre-tetanic synaptic responses. A lower dose of D-APV (1 μM) did not protect i.p.s.ps against the effects of tetani, and a higher dose (100 μM) produced a generalized reduction of e.p.s.ps, i.p.s.ps and population spikes. The involvement of an NMDA receptor in the changes of synaptic potentials is further demonstrated by an increased response to ionophoretically applied NMDA following high frequency stimulation (Fig. 4*a*). The response to ionophoretically applied GABA was decreased progressively by repeated tetanization, being reduced to $< 10\%$ of control values after 4 tetani (see ref. 7). This reduction of GABA sensitivity was observed in all cells in which a reduction of i.p.s.p. amplitude was observed (9 out of 11). In the presence of D-APV, however, the GABA response was not altered by repeated tetani (Fig. 4*b*). After washout of D-APV, further tetanization produced a progressive decrease of the GABA response, a reduction of orthodromically evoked i.p.s.ps, and an increase of e.p.s.ps.

In conclusion, these results demonstrate that the progressive enhancement of excitatory synaptic efficacy and development of epileptiform activity produced by the tetanic stimulation of afferent pathways to CA1 pyramidal cells in the hippocampus

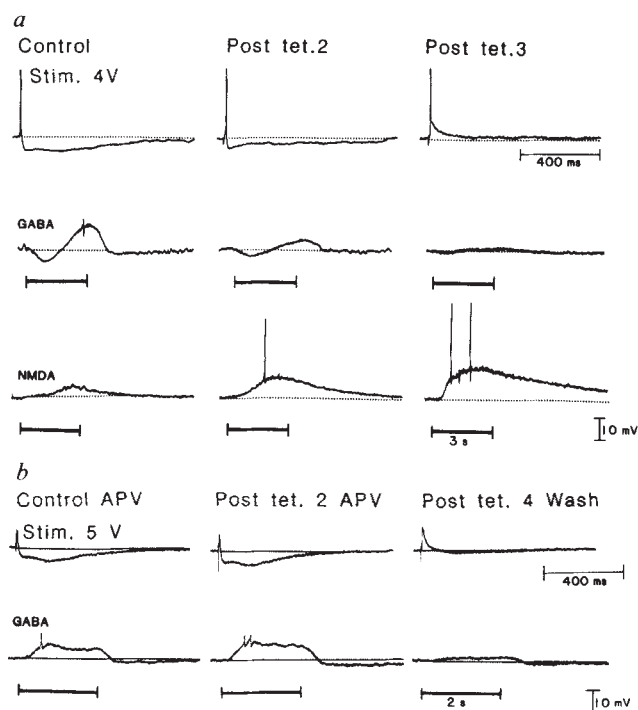


Fig. 4 Effects of tetanic stimulation on ionophoretically applied GABA and NMDA responses. *a*, Progressive fading of the response to ionophoretically applied GABA and the analogous decline of evoked i.p.s.ps. The NMDA response, however, was greatly enhanced following high-frequency stimulation. GABA (ejecting current +11 nA, retaining current -3 nA) and NMDA (ejecting current -22 nA, retaining current +3 nA) were applied to stratum pyramidale through an extracellular 3-barrelled electrode. Resting membrane potential, -73 mV; input resistance, 50 M Ω . The cell membrane potential was held at the resting potential (-73 mV) during NMDA application, and was raised by d.c. current injection to -66 mV during orthodromic stimulation and GABA application. *b*, Control ionophoretic GABA response in solution containing D-APV, after the second tetanus in D-APV and after the fourth tetanus following washout. As for inhibitory synaptic potentials, D-APV suppressed the decrease of the response to ionophoretic GABA application after repeated high-frequency stimulation. Washout and further tetanization produced a depression of both orthodromically evoked i.p.s.ps and GABA response. Action potential amplitudes in *a* and *b* are attenuated. Horizontal bars under GABA and NMDA responses indicate duration of drug ejection. Resting membrane potential, -70 mV; input resistance, 56 M Ω ; GABA ejecting current, +52 nA; retaining current, -8 nA.

is mediated through an alteration of postsynaptic sensitivity to those neurotransmitters believed to be important in both inhibitory neurotransmission and synaptic plasticity. The results also indicate that both the tetanus-induced enhancement of e.p.s.ps and progressive reduction of i.p.s.ps are mediated through postsynaptic NMDA receptor activation, resulting in a much enhanced and highly synchronized orthodromic response which may trigger the genesis of paroxysmal depolarization shifts. The origin of these multiple effects produced by the activation of a single class of postsynaptic receptors may lie in the influx of calcium ions through NMDA receptors activated by the tetanus. Various studies have demonstrated that tetanic stimulation¹³ and NMDA receptor activation¹⁴ lead to an increase of the intracellular calcium concentration, and calcium influx has been demonstrated to reduce the GABA_A-receptor-mediated inward current in bullfrog dorsal root ganglion cells¹⁵.

A probable explanation for the observed alterations in synaptic transmission and postsynaptic sensitivity reported here is that calcium influx through NMDA receptor-ionophores may act at intracellular sites to alter the density or kinetic behaviour

of those receptors involved in synaptic plasticity and transmission in this pathway. The resultant block of recurrent and feedforward i.p.s.ps and the enhancement of a slow, NMDA receptor-mediated component of the e.p.s.p. (normally masked by the shunting effects of i.p.s.ps and channel block by magnesium ions near the resting membrane potential¹⁶) would convert the late biphasic orthodromic i.p.s.p. into a large, long-lasting e.p.s.p. of sufficient magnitude to trigger synchronized epileptiform activity in a large ensemble of pyramidal neurons. This effect would also be facilitated by the increased presynaptic release of excitatory amino acid which is evoked by tetanic stimulation¹⁷. There is considerable evidence that deficits in GABA-mediated neurotransmission may be causal in the genesis of seizures in humans and in animal models of epilepsy^{18,19}. Our results indicate that the synaptic plasticity and epileptiform activity associated with tetanic stimulation of this hippocampal pathway is contingent upon the modulation by NMDA receptors of both excitatory and inhibitory neurotransmission. These synaptic alterations produced by repetitive electrical stimuli may play an important causal role in the kindling model of epilepsy.

We are grateful to Drs M. Galvan, P. Grafe, K. J. Feasey, L. J. Larson-Prior, and J. Behrends for helpful discussions. This research was supported by a DFG grant to M. Galvan through Sonderforschungsbereich 220 (Teilprojekt B2).

Received 6 November 1986; accepted 13 February 1987.

- Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol. Lond.* **334**, 33-46 (1983).
- Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132-137 (1984).
- Wigström, H. & Gustafsson, B. *Neurosci. Lett.* **44**, 327-332 (1984).
- Slater, N. T., Stelzer, A. & Galvan, M. *Neurosci. Lett.* **60**, 25-31 (1985).
- Stasheff, S. F., Bragdon, A. C. & Wilson, W. A. *Brain Res.* **344**, 296-302 (1985).
- Alger, B. E. & Nicoll, R. A. *Brain Res.* **200**, 195-200 (1980).
- Wong, R. K. S. & Watkins, D. J. *J. Neurophysiol.* **48**, 938-951 (1982).
- Andersen, P., Dingledine, R., Gjerstad, L., Langmoen, I. A. & Mosfeldt Laursen, A. J. *Physiol. Lond.* **305**, 279-296 (1980).
- Ben-Ari, Y., Krnjević, K. & Reinhardt, W. C. *Can. J. Physiol. Pharmacol.* **57**, 1462-1466 (1979).
- McCarren, M. & Alger, B. E. *J. Neurophysiol.* **53**, 557-571 (1985).
- Herron, C. E., Williamson, R. & Collingridge, G. L. *Neurosci. Lett.* **61**, 255-260 (1985).
- Dingledine, R., Hynes, M. A. & King, G. L. *J. Physiol. Lond.* **380**, 175-189 (1986).
- Krnjević, K., Morris, M. E. & Ropert, N. *Brain Res.* **374**, 1-11 (1986).
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519-522 (1986).
- Inoue, M., Oomura, Y., Yakushiji, T. & Akaike, N. *Nature* **324**, 156-158 (1986).
- Collingridge, G. L. *Trends pharmac. Sci.* **6**, 407-411 (1985).
- Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496-498 (1982).
- Houser, C. R., Harris, A. B. & Vaughn, J. E. *Brain Res.* **383**, 129-145 (1986).
- Roberts, E. *Adv. Neurol.* **44**, 319-341 (1986).
- Krnjević, K. *Physiol. Rev.* **54**, 418-540 (1974).
- Nicoll, R. A. & Alger, B. E. *Science* **212**, 957-959 (1981).

NMDA receptors of dentate gyrus granule cells participate in synaptic transmission following kindling

Istvan Mody* & Uwe Heinemann*

Department of Neurophysiology, Max Planck Institute for Psychiatry, Am Klopferspitz 18A, 8033 Planegg-Martinsried, FRG

In the mammalian central nervous system, receptors for the excitatory amino-acid neurotransmitters are divided into three subtypes depending on their sensitivity to three specific agonists: kainate, quisqualate and *N*-methyl-D-aspartate (NMDA)¹. The ionophores operated by NMDA are gated by Mg²⁺ in a voltage-dependent manner²⁻⁴ and allow passage of several cations, including Ca²⁺ (refs 5, 6) which may be important in plastic alterations of neuronal excitability. Indeed, specific antagonists of NMDA receptors effectively block spatial learning⁷, long-term potentiation^{7,8} and some animal models of chronic epilepsy^{9,10}. Despite their abundance on central neurons¹¹, NMDA receptors, with a few

* Present addresses: Playfair Neuroscience Unit, Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario M5T 2S8, Canada (I.M.); Institut für Normale und Pathologische Physiologie, Universität Köln, Robert Koch Street 39, 5000 Köln 41, FRG (U.H.).

noteworthy exceptions^{12,13}, do not generally seem to be involved in low-frequency synaptic transmission¹⁴. Here we report for the first time that NMDA receptors of the dentate gyrus, where they do not normally contribute to the generation of synaptic potentials^{7,15,16}, become actively involved in synaptic transmission following long-lasting neuronal changes induced by daily electrical stimulation (kindling)¹⁷ of the amygdala or hippocampal commissures. In contrast to controls, the excitatory postsynaptic potentials (e.p.s.ps) of granule cells in hippocampal slices obtained from kindled animals displayed characteristics typical of an NMDA-receptor-mediated component. The involvement of NMDA receptors in synaptic transmission may underlie the long-lasting changes in neuronal function induced by kindling.

Repetitive stimulation of certain fibre pathways in the central nervous system can induce long-term potentiation (LTP) of excitatory synaptic transmission^{18,19}. If repeated daily, the stimulation will lead, within days or weeks, to the establishment of a chronic epileptic focus that may be active for many months (kindling)¹⁷. Both LTP and kindling share the property of being blocked by NMDA receptor antagonists^{8,10} and in the CA1 region of the hippocampus, kindling produces an enhancement of NMDA-like responses²⁰ suggesting a possible up-regulation of these receptors. In the granule cells of the dentate gyrus, neurons with a high density of NMDA receptors¹¹, the long-lasting biochemical and electrophysiological alterations induced by kindling^{21,22} may also reflect activation of NMDA receptors. Therefore, we have examined the possibility that NMDA receptors are involved in such kindling-induced plastic neuronal changes in the dentate gyrus, where these receptors clearly do not take part in normal synaptic transmission^{7,15,16}. Intracellular recordings were obtained from 20 and 19 granule cells in hippocampal slices from 10 control and 9 kindled animals respectively, using microelectrodes filled with 4 M potassium acetate (resistances: 40–100 M Ω). The input resistance (R_N) of the membrane and the amplitudes of e.p.s.ps were measured at various membrane potentials.

The average resting membrane potential (RMP) and R_N of 20 control granule cells were -68 ± 2 mV (mean $\pm$ s.e.m.) and 37 ± 3 M Ω , respectively. Following suprathreshold stimulation of the lateral perforant pathway or injection of depolarizing current pulses through the electrode, all neurons exhibited action potentials that overshoot the zero potential by 10–20 mV. In all control granule cells, the amplitude of the e.p.s.ps decreased at membrane potentials positive to rest and increased on hyperpolarization of the membrane (Fig. 1a) showing no NMDA-mediated 2-aminophosphonovalerate (APV)-sensitive synaptic component (Fig. 1b; sham-stimulated).

Irrespective of the time elapsed between the last kindling stimulus and the preparation of hippocampal slices (24 hours to 6 weeks), or whether stage IV or V of kindling²³ had been reached, dentate granule cells of kindled animals showed marked differences when compared with controls. Their average RMP was more negative (-73 ± 3 mV), although not significantly different from that of control neurons, but their average R_N (59 ± 5 M Ω) was significantly higher ($P < 0.005$; Student's *t*-test). Subthreshold e.p.s.ps showed a non-linear voltage dependence, in that their amplitude and duration markedly increased at membrane potentials 5–20 mV more positive than rest (Fig. 1a), yet no changes in R_N or membrane time-constant could be detected at these potentials, as measured by hyperpolarizing current pulses delivered before orthodromic stimulation (Fig. 1b, commissural-kindled). The amplitude and voltage dependence of the e.p.s.ps were diminished by addition of 30 μ M APV, indicating the presence of an NMDA-receptor-mediated component in synaptic transmission following kindling (Fig. 1b, commissural-kindled). Burst discharges of action potentials could not be evoked through suprathreshold synaptic activation of kindled granule cells²² in normal perfusate, but frequently ($n = 11$), the action potentials were followed by depolarizing afterpotentials.

Because NMDA-receptor-operated ionophores are gated by Mg^{2+} (refs 2–4), extracellular Mg^{2+} was omitted from the perfusate in several experiments ($n = 15$), revealing further differences between control and kindled granule cells. The passive membrane characteristics and synaptic properties of control neurons were unaltered after 60–90 min in Mg^{2+} -free medium. A 2–5 mV reduction in action potential threshold was observed in 50% of the cells²⁴, but the synaptic potentials of control granule cells were affected less than 5% by the removal of extracellular Mg^{2+} (refs 24 and 25) (Fig. 2a).

In contrast to controls, 25–35 min following washout of extracellular Mg^{2+} , kindled neurons hyperpolarized by 5–8 mV which was accompanied by a 10–20% decrease in R_N , the loss of the voltage dependence of the e.p.s.ps and a strong enhancement of synaptic responses. The stimulus threshold for eliciting e.p.s.ps in kindled neurons was lowered by 20–30% in Mg^{2+} -free medium, as would be expected after the removal of the Mg^{2+} block from the NMDA receptors actively involved in synaptic transmission. Suprathreshold stimulation in this medium induced bursts of action potentials (Fig. 2b) correlated with multiple population spikes in extracellular recordings, not normally encountered in the control dentate gyrus^{24,25} (Fig. 2a). The synaptically evoked action potential bursts in Mg^{2+} -free medium and the large, long duration depolarizing afterpotentials, both in normal and Mg^{2+} -free perfusate, were reversibly (not shown) antagonized by addition of 30 μ M APV (Fig. 2b, $n = 8$).

The voltage dependence of the e.p.s.ps, their sensitivity to extracellular Mg^{2+} and APV encountered in kindled granule cells, are regular characteristics of neurons with NMDA receptors actively involved in synaptic transmission. The presence of an NMDA-receptor-mediated synaptic component following kindling may produce long-lasting alterations in granule cell function through activation of Ca^{2+} -dependent mechanisms because of the large Ca^{2+} permeability that distinguishes the ionophores activated by NMDA from all other amino-acid receptor subtypes^{5,6}. An increased steady-state Ca^{2+} uptake into kindled granule cells mediated by active NMDA receptors is therefore possible²⁰. Such changes in cellular Ca^{2+} homeostasis have already been shown in kindled granule cells through specific loss of an intraneuronal Ca^{2+} -binding protein²¹.

In addition, tonically active NMDA receptors may be responsible for the altered membrane characteristics of kindled granule cells, particularly the increased R_N , because iontophoresis of NMDA onto control granule cells causes similar increases in R_N ²⁶. In the present study, APV usually reduced the R_N of kindled neurons but not that of controls, indicating that NMDA receptors are tonically activated. If this proves to be so, the effects of other neurotransmitters, such as noradrenaline, that interact with NMDA receptors in the dentate gyrus²⁵ will strongly depend on the state of activation of these receptors. Furthermore, the responses of kindled granule cells to temporally associated synaptic inputs will be dramatically affected by the voltage dependence of the e.p.s.ps and the increased neuronal R_N . Subthreshold synaptic potentials may thus become suprathreshold during kindling and could lead to the establishment of the circuitry necessary for generalized convulsions, as shown in the CA1 region of hippocampal slices where NMDA receptor antagonists effectively attenuate the epileptiform activity induced by convulsants²⁷ or massive repetitive electrical stimulation²⁸.

In summary, our findings reveal that the contribution of NMDA receptors to synaptic transmission in a system where responses mediated by these receptors are not normally expressed may constitute the basis for the long-lasting changes in neuronal function following kindling. The synaptic NMDA component in the kindled dentate gyrus may result from the following mechanisms, or combinations of them: (1) progressive activation of dormant NMDA receptors; (2) increase in the number of NMDA receptors; (3) changes in dendritic spine

Fig. 1 Demonstration of the voltage-dependent APV-sensitive synaptic NMDA component in granule cells following kindling. *a*, Normalized amplitude of the intracellularly recorded e.p.s.ps as a function of the membrane potential in control (○; $n = 19$) and kindled (●; $n = 17$) neurons. The average amplitude (range: 5–15 mV) of the e.p.s.ps at RMP ± 2.5 mV was considered to be 100% (error bars, s.e.m.). In contrast to controls, the amplitude of e.p.s.ps increased at depolarized membrane potentials in kindled granule cells (*, $P < 0.005$, Student's two tailed *t*-test). *b*, Sensitivity of synaptic transmission to membrane potential and APV in a granule cell ($R_N = 54$ M Ω) from a commissural-kindled animal and in a granule cell ($R_N = 35$ M Ω) from a sham-stimulated control. The traces represent intracellular voltage deflections produced by injections of -0.2 nA current pulses of 60 ms duration followed by stimulation of the lateral perforant path. In both cells, the responses were taken at rest (RMP) and at membrane potentials 10 mV depolarized from rest. The amplitude and duration of the e.p.s.p. of the kindled granule cell increased on depolarization of the membrane in sharp contrast to the control neuron. Perfusion of APV (30 μ M) reduced the amplitude of the kindled granule cell e.p.s.p. at both membrane potentials, revealing a clear NMDA component of the synaptic potential. Right panel, digital subtractions (LSI 11/73 computer) of the responses in 30 μ M APV from those in control perfusate demonstrates the synaptic component blocked by APV, that is, the contribution of NMDA receptors to kindled synaptic transmission. There is no such contribution to the e.p.s.ps of the sham-stimulated granule cell at either membrane potential.

Methods. Male Wistar rats (180–220 g) were stereotactically implanted with stimulation electrodes under pentobarbital anaesthesia (70 mg per kg), either into the midline hippocampal commissural pathway or the amygdaloid nucleus. Starting on days 7–10 after surgery, the animals were stimulated daily¹⁷ with a 1-s, 60-Hz stimulus train of 100–150 μ A until kindling stage IV or V (commissural) and stage V (amygdala) was reached²³. Sham-stimulated animals were handled in all aspects like kindled ones, but received only 0–3 stimuli. At various times following the last stimulation, kindled and sham-stimulated animals were killed under ether anaesthesia, their brains removed and the hippocampi were dissected free. Hippocampal slices (425 μ m thick) were maintained *in vitro* at 35.5 °C and were perfused with oxygenated artificial cerebrospinal fluid containing (mM): NaCl 126, KCl 5, NaH_2PO_4 1.25, NaHCO_3 26, CaCl_2 2, MgSO_4 2 or 0, and D-glucose 10. When used, DL-APV (Cambridge Research) was added to yield a final concentration of 30 μ M. Thus, the D-APV concentration effective for blocking NMDA receptors was probably around 15 μ M. Bipolar, glass-insulated platinum wire electrodes were placed at the border of the subiculum and the dentate gyrus for stimulation of the lateral perforant path.

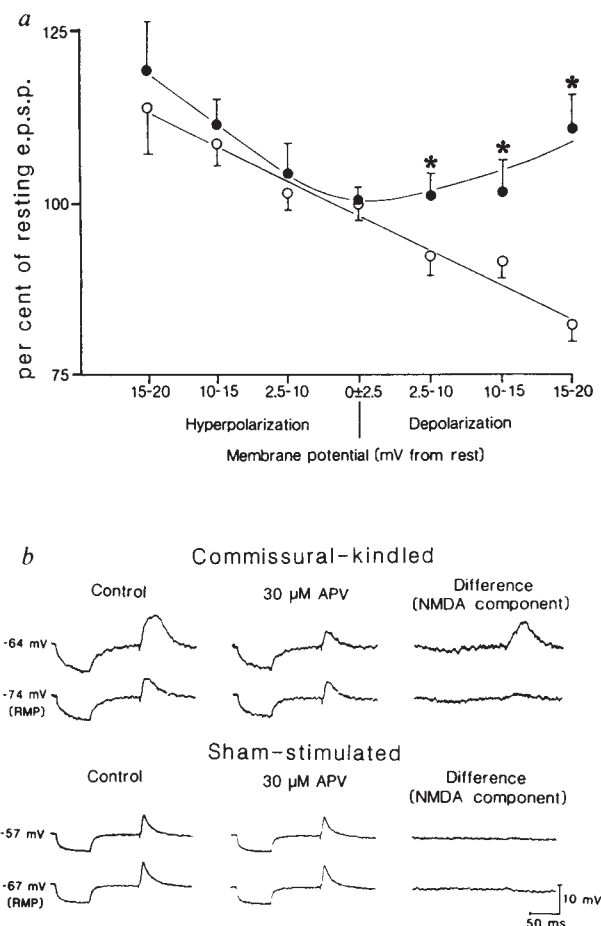
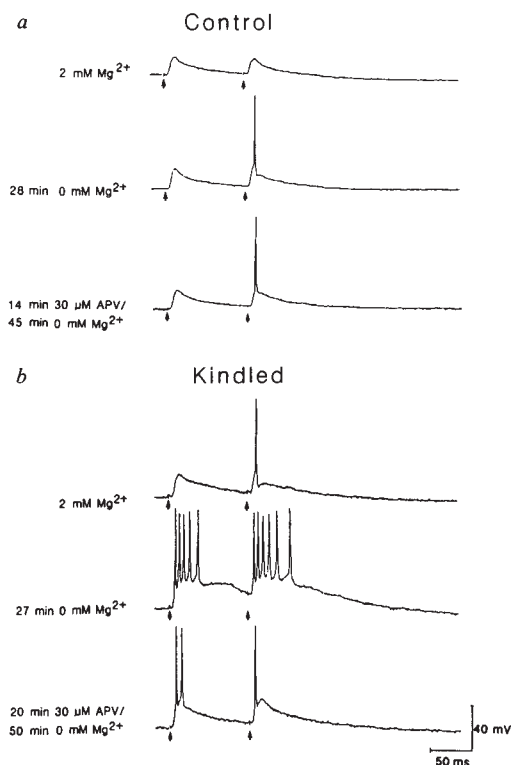


Fig. 2 Effect of omission of extracellular Mg^{2+} from the perfusate on synaptic properties of control and kindled granule cells. *a*, Intracellular responses to paired-pulse synaptic stimulation (arrows) of a dentate granule cell in a hippocampal slice from a sham-stimulated control animal. The responses were taken at the times indicated following washout of Mg^{2+} . The absence of extracellular Mg^{2+} did not alter the RMP (-71 mV) and the R_N (42 M Ω) of the neuron. Addition of 30 μ M APV to the Mg^{2+} -free perfusate was ineffective, indicating that NMDA receptors do not participate in synaptic transmission in spite of the removal of the Mg^{2+} -blockade from these receptors. *b*, Effects of omission of extracellular Mg^{2+} and perfusion of 30 μ M APV on the e.p.s.ps of a kindled granule cell. The hippocampal slice was prepared 6 weeks following the last stimulation of an amygdala-kindled animal. All responses are taken at the same membrane potential (-72 mV), although the absence of extracellular Mg^{2+} hyperpolarized the neuron to -78 mV and decreased its R_N from 69 to 60 M Ω . The synaptic responses to paired-pulse stimulation (arrows) of the lateral perforant path were obtained at the times indicated. The absence of extracellular Mg^{2+} (for longer than 20 min) enhanced the synaptic responses, producing bursts of action potential firing. A 20-min perfusion of 30 μ M APV reversibly antagonized the burst firing and the large depolarizing afterpotentials in Mg^{2+} -free medium (data not shown).

morphology resulting in the exposure of NMDA receptors to the neurotransmitter; (4) enhanced neurotransmitter release from perforant path terminals causing an 'overspill' of the endogenous transmitter to reach extrasynaptically located NMDA receptors. Whether or not similar mechanisms function in other forms of drug- or stimulus-induced plastic neuronal alterations, including epileptiform activity, has yet to be determined.

This research was supported by DFG-grants from the Deutsche Forschungsgemeinschaft and the SFB (Sonderforschungsbereich 220) to U.H., and MRC of Canada and I. W. Killam Fellowships to I.M. We thank Dr J. J. Hablitz for critical comments on early drafts of this manuscript and Beate Muffler for technical assistance.

Received 9 December 1986; accepted 12 February 1987.

1. Watkins, J. C. & Evans, R. H. A. *Rev. Pharmac. Tox.* **21**, 165-204 (1981).
2. Adult, B., Evans, R. H., Francis, A. A., Oakes, D. J. & Watkins, J. C. *J. Physiol., Lond.* **307**, 413-428 (1980).
3. Nowak, L., Bregestovski, P., Asher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).

4. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261-263 (1984).
5. MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519-522 (1986).
6. Lambert, J. D. C. & Heinemann, U. in *Calcium Electrogenesis and Neuronal Functioning* (eds Heinemann, U. et al.) 279-290 (Springer, Heidelberg, 1986).
7. Morris, R. G. M., Anderson, E., Lynch, G. S. & Baudry, M. *Nature* **319**, 774-776 (1986).
8. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol., Lond.* **334**, 33-46 (1983).
9. Croucher, M. J., Collins, J. F. & Meldrum, B. S. *Science* **216**, 899-901 (1982).
10. Peterson, D. W., Collins, J. F. & Bradford, H. F. *Brain Res.* **311**, 176-180 (1984).
11. Monaghan, D. T. & Cotman, C. W. *J. Neurosci.* **5**, 2909-2919 (1985).
12. Thomson, A. M., West, D. C. & Lodge, D. *Nature* **313**, 479-481 (1985).
13. Salt, T. E. *Nature* **322**, 263-265 (1986).
14. Watkins, J. C. *Trends pharmac. Sci.* **5**, 373-376 (1984).
15. Crunelli, V., Forda, S. & Kelly, J. S. *J. Physiol., Lond.* **341**, 627-640 (1983).
16. Mody, I. & Heinemann, U. *Neurosci. Lett.* **69**, 137-142 (1986).
17. Goddard, G. V. *Nature* **214**, 1020-1021 (1967).
18. Bliss, T. V. P. & Lomo, T. J. *Physiol., Lond.* **232**, 334-356 (1973).
19. Bliss, T. V. P. & Gardner-Medwin, A. R. *J. Physiol., Lond.* **232**, 357-374 (1973).
20. Wadman, W. J., Heinemann, U., Konnerth, A. & Neuhäus, S. *Exp Brain Res.* **57**, 404-407 (1985).
21. Miller, J. J. & Baimbridge, K. G. *Brain Res.* **278**, 322-326 (1983).
22. Oliver, M. W. & Miller, J. J. *Exp Brain Res.* **57**, 443-447 (1985).
23. Racine, R. J. *Electroenceph. clin. Neurophysiol.* **32**, 281-294 (1972).
24. Mody, I., Lambert, J. D. C. & Heinemann, U. *J. Neurophysiol.* (in the press).
25. Stanton, P. K., Jones, R. S. G., Mody, I. & Heinemann, U. *Epilepsy Res.* **1**, 53-62 (1987).
26. Crunelli, V. & Mayer, M. L. *Brain Res.* **311**, 392-396 (1984).
27. Herron, C. E., Williamson, R. & Collingridge, G. L. *Neurosci. Lett.* **61**, 255-260 (1985).
28. Slater, N. T., Stelzer, A. & Galvan, M. *Neurosci. Lett.* **60**, 25-31 (1985).

Synapsin I bundles F-actin in a phosphorylation-dependent manner

Martin Böhler & Paul Greengard

Laboratory of Molecular and Cellular Neuroscience,
The Rockefeller University, 1230 York Avenue, New York,
New York 10021-6399, USA

Synapsin I is a neuron-specific phosphoprotein localized to the cytoplasmic surface of synaptic vesicles^{1,2}. This phosphoprotein is a major substrate for cyclic AMP-dependent and calcium/calmodulin-dependent protein kinases³⁻⁵. Its state of phosphorylation can be altered both *in vivo* and *in vitro* by a variety of physiological and pharmacological manipulations known to affect synaptic function^{6,7}. Recent direct evidence suggests that it may be involved in the regulation of neurotransmitter release from the nerve terminal⁸. In the nerve terminal, synaptic vesicles are embedded in a cytoskeletal network, consisting in part of actin⁹⁻¹¹. We report here the ability of the dephospho-form of synapsin I to bundle F-actin. This bundling activity is reduced when synapsin I is phosphorylated by cAMP-dependent protein kinase and virtually abolished when it is phosphorylated by calcium/calmodulin-dependent protein kinase II or by both kinases. These results, demonstrating an interaction of synapsin I with actin *in vitro*, support the possibility that synapsin I is involved in clustering of synaptic vesicles at the presynaptic terminal and that the phosphorylation of synapsin I may be involved in regulating the translocation of synaptic vesicles to their sites of release.

Negative-stain electron microscopy revealed that when F-actin was incubated with dephospho-synapsin I, long F-actin bundles were formed (compare Fig. 1a, b and inset). This suggested that synapsin I can bind to the sides of actin filaments. Synapsin I has a collagenase-resistant head region and a collagenase-sensitive tail region and is phosphorylated in the head region by cAMP-dependent protein kinase or calcium/calmodulin-dependent protein kinase I and in the tail region by calcium/calmodulin-dependent protein kinase II. When F-actin samples were incubated with synapsin I which had been phosphorylated in both the head and tail regions, virtually no actin bundles were observed (Fig. 1c).

To quantify the bundling activity of synapsin I, a light-scattering assay was used (Fig. 2a). Dephospho-synapsin I was found to be effective even at low concentrations in increasing light scattering of F-actin solutions. Synapsin I phosphorylated in the head region increased light scattering, but less effectively

than did dephospho-synapsin I. Synapsin I phosphorylated in the tail region, or in both the head and tail regions, had little or no effect on the light scattering of F-actin solutions. When control experiments were performed with synapsin I which had been subjected to pseudo-phosphorylation by omitting ATP, the results obtained were identical to those obtained with dephospho-synapsin I (data not shown). When the salt concentration of the incubation medium was reduced from 100 mM KCl to 50 mM KCl, less dephospho-synapsin I was needed to increase light scattering to the same extent. Under these low salt conditions, however, the differences between the different phosphorylation forms of synapsin I in their ability to increase light scattering of F-actin solutions were almost abolished (data not shown).

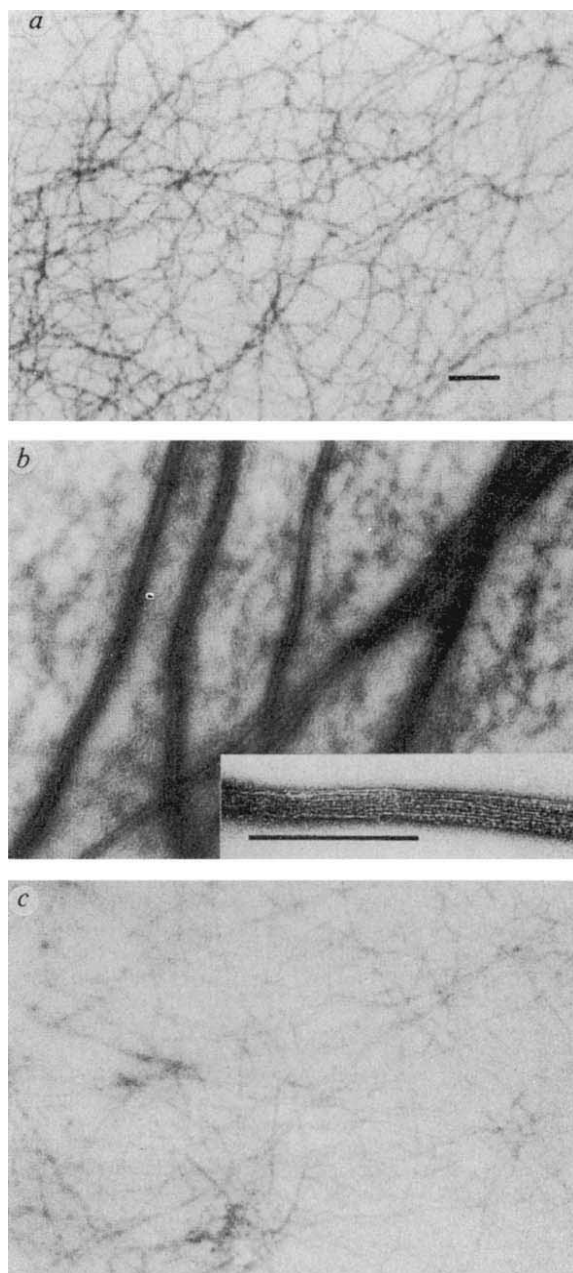
Synapsin I binding to F-actin was directly demonstrated by low-speed sedimentation experiments. Under low-speed conditions actin bundles sediment, whereas non-bundled F-actin does not. Dephospho-synapsin I was found to be the most potent of the four forms of synapsin I in bundling actin (Fig. 2b). In addition, the amount of synapsin I found in the pellets was proportional to that of actin. These results are consistent with the electron microscope and light-scattering experiments.

To characterize the interaction of synapsin I with actin further, we investigated synapsin I for possible effects on actin polymerization, using pyrene-labelled actin as a probe. Dephospho-synapsin I inhibited actin polymerization, synapsin I phosphorylated in the head region was less effective, and synapsin I phosphorylated in the tail region, or in both the head and tail regions, had no effect on polymerization (Fig. 3a). Inhibitory activity of dephospho-synapsin I was also observed when polymerization was nucleated by the addition of F-actin seeds (Fig. 3b), indicating that synapsin I interfered with elongation rather than with nucleation. The mechanism by which synapsin I reduces the rate of actin polymerization is not clear, but there are several reasons for believing that this effect may be secondary to the bundling of F-actin. First, it is reminiscent of findings with protein 4.9 and high magnesium concentrations, both of which bundle F-actin¹². Second, reducing the synapsin I concentration to slightly below that necessary for significant bundling of F-actin abolished the effect on actin elongation. Third, synapsin I had no detectable effect on the critical actin concentration necessary for actin polymerization, as measured by high shear viscosity (data not shown). Finally, synapsin I had no effect on the ratio of F-actin to G-actin, as indicated by equal pelletable amounts of F-actin in the presence and absence of synapsin I (data not shown).

Bundling of F-actin by dephospho-synapsin I suggests either that dephospho-synapsin I has more than one site which can

Fig. 1 Effect of synapsin I on actin filament bundle formation as analysed by electron microscopy. *a*, F-actin; *b*, F-actin plus dephospho-synapsin I; *c*, F-actin plus synapsin I which had been phosphorylated in the head region by cAMP-dependent protein kinase (0.75 mol phosphate per mol synapsin I) and in the tail region by calcium/calmodulin-dependent protein kinase II (1.75 mol phosphate per mol synapsin I). Inset in *b* shows bundled actin filaments at a higher magnification.

Methods. G-actin (5 μ M) in the absence or presence of synapsin I (1 μ M) was polymerized for 2 h at 4 °C in 70 mM KCl, 26 mM NaCl, 2 mM MgCl₂, 3.4 mM Tris/HCl (pH 8.0), 15 mM HEPES (pH 7.4), 0.5 mM 2-mercaptoethanol, 35 μ M CaCl₂, 0.1 mM EGTA. Samples were applied to carbon-coated grids, rinsed with 0.1 M ammonium acetate, negatively stained with 1% uranyl acetate and examined in the electron microscope. Calibration bars, 0.5 μ m. Actin was prepared from an acetone powder of rabbit skeletal muscle²⁶ and gel filtered on Sephadex G-150²⁷ as described. Synapsin I was purified from bovine brain by a modification of a published procedure¹³. Briefly, cleaned calf cerebral cortex was homogenized in 0.32 M sucrose, 4 mM EDTA, 1 mM EGTA, 0.5 mM 2-mercaptoethanol, pH 7.4 and a protease inhibitor mixture consisting of 0.1 mM phenylmethylsulphonylfluoride, 10 μ g ml⁻¹ leupeptin, 2 μ g ml⁻¹ pepstatin, 10 μ g ml⁻¹ chymostatin and 10 μ g ml⁻¹ antipain. After centrifugation at 8,600g for 45 min, the pellet was extracted twice with 0.15 M NaCl, 1% Nonidet-P40, 2 mM EDTA, 0.5 mM 2-mercaptoethanol, 3 mM HEPES (pH 7.4) (final concentrations) and the protease inhibitor mixture. The extract was diluted fivefold with 2 mM Tris/HCl (pH 8.0) and batch adsorbed to carboxymethylcellulose, pre-equilibrated in 10 mM Tris/HCl (pH 8.0), 0.1 mM EGTA, 1 mM EDTA, 0.5 mM 2-mercaptoethanol and 3 mM sodium azide. The CM-cellulose was packed into a column and developed with a salt gradient (20–200 mM NaCl in the equilibration buffer). The pooled synapsin I fractions were diluted threefold with water and loaded onto a hydroxylapatite column equilibrated with 10 mM NaPO₄ (pH 7.4), 2 mM EDTA, 0.1 mM EGTA and 0.5 mM 2-mercaptoethanol. Synapsin I was eluted with a 0–0.3 M NaCl gradient in the equilibration buffer, precipitated with 35% ammonium sulphate, redissolved in a solution containing 0.2 M NaCl, 25 mM Tris/HCl (pH 7.5), 1 mM EDTA, 1 mM EGTA, 5 mM 2-mercaptoethanol, 3 mM sodium azide, and chromatographed over a Sephacryl S-200 gel filtration column. Purified dephospho-synapsin I was phosphorylated to near stoichiometry and repurified by adsorption to carboxymethyl-cellulose as described¹³, except that Triton X-100 was omitted from the head region phosphorylation mixtures. Purified kinases and calmodulin were prepared as described^{28–30} and were gifts of Mrs A. Horiuchi, Dr Y. Lai and Dr A. C. Nairn. A trace amount of [γ -³²P]ATP was added to the reaction mixtures for determination of the stoichiometry of phosphorylation. Incorporation of phosphate into head and tail regions was analysed by one-dimensional proteolytic phosphopeptide mapping.



bind to F-actin, or that it has a single binding site and achieves bundling by self-association. As a step towards understanding the role of synapsin I phosphorylation in regulating F-actin bundling, we examined the effect of phosphorylation of synapsin I on its binding to F-actin. Non-linear regression analysis of the binding data (Fig. 4), assuming a one-binding-site model, showed no significant effect of phosphorylation on the dissociation constant (K_d), the mean K_d value for the four forms of synapsin I being 1.7 μ M and 1.5 μ M in two experiments. Because of the likelihood of overestimation of free synapsin I (see Fig. 4 legend), the K_d values are maximum values. At saturation, one mole of dephospho-synapsin I was bound per 6.8 moles of actin. Interestingly, we found a marked reduction of the maximum binding (B_{max}) for the phosphoforms of synapsin I (1:10.6 to 1:7 for synapsin I phosphorylated in the head region, 1:11.2 for synapsin I phosphorylated in the tail region and 1:11.5 for synapsin I phosphorylated in both the head and tail regions). The fact that phosphorylation of synapsin I altered the B_{max} rather than the K_d is compatible with a self-association model. However, we have not been able to detect any phosphory-

lation-dependent changes in self-association of synapsin I using cross-linking procedures.

Evidence of an interaction between synapsin I and actin was obtained using two additional techniques (data not shown). First, it was found that synapsin I affects the low shear viscosity of actin solutions in a phosphorylation-dependent fashion. Second under non-polymerizing conditions synapsin I was found to interact with G-actin in a phosphorylation-independent fashion leading to the formation of bundled aggregates (data not shown); it is not known, however, whether synapsin I interacts with G-actin under conditions in which F-actin is also present.

The detailed mechanism by which phosphorylation of synapsin I affects its ability to bundle F-actin remains to be determined. It will also be important to determine whether the interaction of synapsin I with actin is of physiological significance. Synapsin I has a very basic isoelectric point of ~ 11.5 (ref. 13) and it is well established that a number of basic macromolecules such as polylysine, histones, lysozyme and nerve growth factor, can bundle actin^{14–16}. But the fact that phosphorylation of synap-

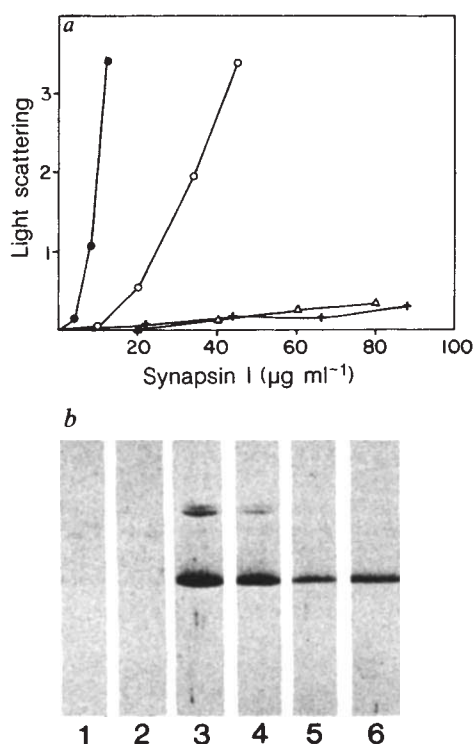


Fig. 2 Effect of synapsin I on actin filament bundle formation determined by *a*, light scattering and *b*, low-speed centrifugation. *a*, F-actin was incubated with dephospho-synapsin I (●), or synapsin I phosphorylated in the head region (○), in the tail region (△), or in both (+). *b*, Coomassie blue stained gel of pellets obtained after low speed centrifugation. Lane 1, actin alone; lane 2, dephospho-synapsin I alone; lane 3, actin plus dephospho-synapsin I; lane 4, 5 and 6, actin plus synapsin I phosphorylated in the head region (lane 4), the tail region (lane 5) or both (lane 6).

Methods. In *a*, aliquots (1 ml) of F-actin (0.06 mg ml^{-1}) in 100 mM KCl, 2 mM MgCl_2 , 1 mM ATP, 10 mM HEPES (pH 7.4) and 0.5 mM 2-mercaptoethanol were incubated with the indicated final concentrations of dephospho-synapsin I, synapsin I phosphorylated in the head region (0.8 mol phosphate per mol synapsin I), synapsin I phosphorylated in the tail region (1.8 mol phosphate per mol synapsin I) or synapsin I phosphorylated in both the head region (0.75 mol phosphate per mol synapsin I) and tail region (1.75 mol phosphate per mol). Stock solutions of synapsin I were prepared in 175 mM NaCl, 0.1 mM EGTA, 25 mM Tris/HCl (pH 8.0), at comparable protein concentrations. Light scattering was measured as described^{31,32} at an angle of 90° in a Perkin-Elmer 650-40 fluorescence spectrophotometer with excitation and emission wavelengths set at 400 nm and slits set at 2 nm. An increase of one unit in scattering represents a 20-fold increase over the light scattering of F-actin alone. In *b*, G-actin ($5 \mu\text{M}$) or synapsin I ($1 \mu\text{M}$), or both, were incubated under polymerization conditions for 3 h at 4°C in 70 mM KCl, 26 mM NaCl, 2 mM MgCl_2 , 0.2 mM ATP, 0.5 mM 2-mercaptoethanol, 15 μM EGTA, 35.6 μM CaCl_2 , 4.1 mM Tris/HCl (pH 8.0) and 15 mM HEPES (pH 7.4). 1, actin alone; 2, dephosphosynapsin I alone; 3, actin plus dephospho-synapsin I; in lane 4, synapsin I phosphorylated in the head region (0.8 mol phosphate per mol synapsin I) was used; in lane 5, synapsin I phosphorylated in the tail region (1.8 mol phosphate per mol synapsin I); in lane 6, synapsin I phosphorylated in both the head region (0.75 mol phosphate per mol synapsin I) and tail region (1.75 mol phosphate per mol synapsin I). Pellets obtained upon centrifugation ($10,000g$ for 10 min) were analysed by Coomassie blue staining of SDS-polyacrylamide gels.

sin I appears to abolish its ability to bundle F-actin, although it has only a slight effect on its isoelectric point, supports the possibility that this interaction might be of physiological significance. Moreover, synapsin I is a highly surfactant molecule (M.Ho *et al.*, manuscript in preparation) and it is possible that the interaction with actin is attributable to this property. The

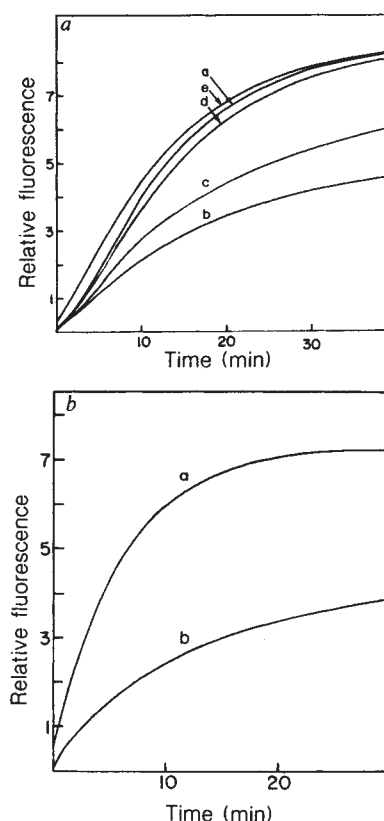


Fig. 3 Effect of synapsin I on actin polymerization measured with a pyrene-labelled actin probe. Polymerization was assayed *a*, in the absence or *b*, presence of F-actin seeds ($12.4 \mu\text{g ml}^{-1}$ sheared F-actin filaments). Samples contained $3.7 \mu\text{M}$ actin (5% pyrene-labelled) in 1 ml of 50 mM KCl, 35 mM NaCl, 2 mM MgCl_2 , 1 mM ATP, 23 μM CaCl_2 , 1 mM EGTA, 0.5 mM dithiothreitol, 5 mM Tris/HCl (pH 8.0), 20 mM HEPES (pH 7.4) and $0.6 \mu\text{M}$ of the indicated form of synapsin I. At time zero, G-actin was added and the polymerization that followed was monitored as an increase in the fluorescence of the sample. *a*, Actin alone; *b*, actin plus dephospho-synapsin I; *c*, actin plus synapsin I phosphorylated in the head region (0.8 mol phosphate per mol synapsin I); *d*, actin plus synapsin I phosphorylated in the tail region (1.8 mol phosphate per mol synapsin I); and *e*, actin plus synapsin I phosphorylated in both the head region (0.75 mol phosphate per mol synapsin I) and tail region (1.75 mol phosphate per mol synapsin I). Actin was labelled with pyrene as described^{33,34}. All measurements were made in a Perkin-Elmer 650-40 recording fluorescence spectrophotometer with the wavelengths for excitation and emission set at 365 and 407 nm, respectively. The excitation and emission slit widths were set at 2 and 10 nm, respectively.

fact that phosphorylation has no detectable effect on its surface activity argues against this possibility, however.

It has been suggested that synapsin I might link synaptic vesicles to the cytoskeleton^{1,17}. There have been several recent reports that synapsin I can interact *in vitro* with various cytoskeletal components, namely spectrin, microtubules and neurofilaments¹⁸⁻²⁰. Unfortunately, it was not determined whether any of those interactions were affected by the state of phosphorylation of synapsin I.

Synapsin I has been proposed as a brain analogue of erythrocyte protein 4.1 (ref. 18). Interestingly, phosphorylation of erythrocyte protein 4.1 has been reported to decrease its affinity for spectrin²¹. Erythrocyte protein 4.1 does not exert any effect on actin *per se*²², but does so in the presence of spectrin^{23,24}. Future studies of the ternary complex synapsin I-spectrin-actin should take into account the results reported here on the binary interactions between synapsin I and actin.

Whether synapsin I interacts with any cytoskeletal components *in vivo*, and whether such interactions are modified by

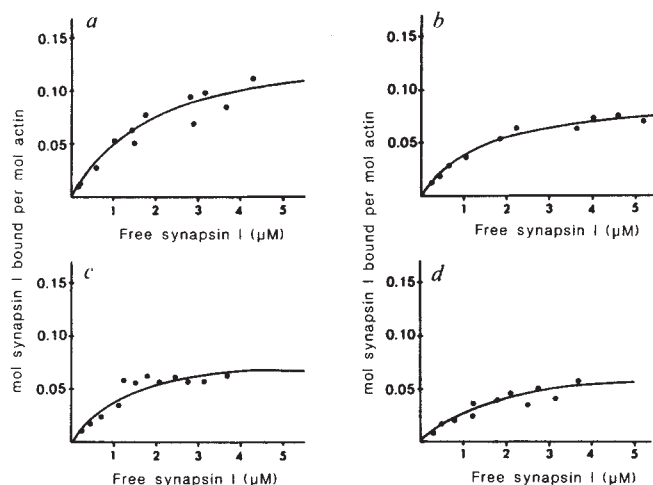


Fig. 4 Binding of synapsin I to F-actin. F-actin (4 μM) was incubated with various amounts of: *a*, dephospho-synapsin I; *b*, synapsin I phosphorylated in the head region (0.8 mol phosphate per mol synapsin I); *c*, synapsin I phosphorylated in the tail region (1.8 mol phosphate per mol synapsin I); and *d*, synapsin I phosphorylated in both head region (0.75 mol phosphate per mol synapsin I) and tail region (1.75 mol phosphate per mol synapsin I).

Methods. The synapsin I solution was clarified by prespinning in a Beckman TL-100 ultracentrifuge at 200,000g for 15 min. F-actin plus synapsin I or synapsin I alone were incubated in 50 μl of 70 mM NaCl, 60 mM KCl, 1.2 mM MgCl₂, 0.6 mM ATP, 0.3 mM 2-mercaptoethanol, 84 μM CaCl₂, 40 μM EGTA, 14.2 mM Tris/HCl and 9.9 mM HEPES, pH 7.4 for 2 h on ice. (Synapsin I was added last using pipette tips that were pre-coated with synapsin I. Aliquots added with this pipette tip were analysed for the amounts of synapsin I pipetted.) Following incubation, free synapsin I was separated from bound synapsin I by centrifugation in a Ti 42.2 rotor at 42,000 r.p.m. for 20 min. Supernatants were carefully aspirated and pellets resuspended in 40 μl of 4.7 M NH₄Cl. The amount of synapsin I pelleted was quantified by a dot-immunobinding assay³⁵ after appropriate dilution into 'spot-buffer' (150 mM NaCl, 10 mM Na-phosphate, pH 7.4, 1.4% Triton X-100 (v/v)). The amount of actin pelleted was quantified by densitometry of Coomassie blue stained gels. The free synapsin I concentration was calculated by subtracting the amount pelleted from the total amount added. (It was assumed that all the F-actin present in the assay mixture pelleted and that synapsin I did not interact with G-actin. Synapsin I adsorbed to a significant extent to tubewalls. This tubewall-adsorbed synapsin I was included in our calculations of free synapsin I.) The binding data were analysed for determination of K_d and B_{max} values by non-linear regression analysis (one binding site model) by the computer program Recept³⁶.

phosphorylation, are important topics for future investigation. Microtubules and neurofilaments are thought not to extend into the vesicle area of the presynaptic terminal^{10,25}. In contrast, actin is found in the vesicle area (refs 9–11; T. Reese, personal communication). Therefore, synaptic vesicle-synapsin I-actin interactions, regulated by phosphorylation of synapsin I, may be important in the regulation of vesicle movement and/or neurotransmitter release. It is of interest, in relation to the results of the present study, that injected dephospho-synapsin I, but not synapsin I phosphorylated in the tail region, inhibits neurotransmitter release from the squid giant synapse terminal⁸.

We thank Mrs Atsuko Horiuchi, Dr Yvonne Lai and Dr Angus Nairn for their generous gift of protein kinases and calmodulin, Drs Fabio Benfenati and Vincenzo Guardabasso for the use of their computer program in analysing the binding data and Drs Fabio Benfenati and Flavia Valtorta for helpful discussions. This work was supported by a grant from the US PHS (MH-39327) to P.G. M.B. was the recipient of a fellowship from the Swiss National Science Foundation.

Received 19 December 1986; accepted 5 February 1987.

1. DeCamilli, P., Harris, J. M. Jr, Huttner, W. B. & Greengard, P. *J. Cell Biol.* **96**, 1355–1373 (1983).

2. Huttner, W. B., Schiebler, W., Greengard, P. & DeCamilli, P. *J. Cell Biol.* **96**, 1374–1388 (1983).
3. Johnson, E. M., Ueda, T., Maeno, H. & Greengard, P. *J. Biol. Chem.* **247**, 5650–5662 (1972).
4. Huttner, W. B. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5402–5406 (1979).
5. Kennedy, M. B. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1293–1297 (1981).
6. Dolphin, A. C. & Greengard, P. *Nature* **289**, 76–79 (1981).
7. Nestler, E. J. & Greengard, P. in *Protein Phosphorylation in the Nervous System* (Wiley, New York, 1984).
8. Llinas, R., McGuinness, T. L., Leonard, C. S., Sugimori, M. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3035–3039 (1985).
9. Toh, B. H. et al. *Nature* **264**, 648–650 (1976).
10. Walker, J. H. et al. *Eur. J. Cell Biol.* **38**, 123–133 (1985).
11. LeBeux, Y. & Willemot, J. *Cell Tiss. Res.* **160**, 37–68 (1975).
12. Siegel, D. L. & Branton, D. *J. Cell Biol.* **100**, 775–785 (1985).
13. Schiebler, W., Jahn, R., Doucet, J. P., Rothlein, J. & Greengard, P. *J. Biol. Chem.* **261**, 8383–8390 (1986).
14. Calissano, P., Monaco, G., Castellani, G., Mercanti, D. & Levi, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2210–2214 (1978).
15. Griffith, L. M. & Pollard, T. D. *J. Biol. Chem.* **257**, 9135–9142 (1982).
16. Fowler, W. E. & Aebi, U. *J. Cell Biol.* **93**, 452–458 (1982).
17. DeCamilli, P. & Greengard, P. *Biochem. Pharmacol.* **35**, 4349–4357 (1986).
18. Baines, A. J. & Bennett, V. *Nature* **315**, 410–413 (1985).
19. Baines, A. J. & Bennett, V. *Nature* **319**, 145–147 (1986).
20. Goldenring, J. R. et al. *J. Biol. Chem.* **263**, 8495–8504 (1988).
21. Eder, P. S., Soong, C. J. & Tao, M. *Biochemistry* **25**, 1764–1770 (1986).
22. Pinder, J. C., Ohanian, V. & Gratzner, W. B. *FEBS Lett.* **169**, 161–164 (1984).
23. Ungewickell, E., Bennett, P. M., Calvert, R., Ohanian, V. & Gratzner, W. B. *Nature* **280**, 811–814 (1979).
24. Fowler, V. & Taylor, D. L. *J. Cell Biol.* **85**, 361–376 (1980).
25. Roots, B. I. *Science* **221**, 971–972 (1983).
26. Spudich, J. A. & Watt, S. J. *J. Biol. Chem.* **246**, 4866–4871 (1971).
27. MacLean-Fletcher, S. & Pollard, T. D. *Biochem. biophys. Res. Commun.* **96**, 18–27 (1980).
28. Kaczmarek, L. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7487–7491 (1980).
29. McGuinness, T. L., Lai, Y. & Greengard, P. *J. Biol. Chem.* **260**, 1696–1704 (1985).
30. Watterson, D. M., Harrelson, W. G., Keller, P. M., Sharief, F. & Vanaman, T. C. *J. Biol. Chem.* **251**, 4501–4513 (1976).
31. Ando, T. & Scales, D. J. *Biol. Chem.* **260**, 2321–2327 (1985).
32. Wilkins, J. A. & Lin, S. J. *J. Cell Biol.* **102**, 1085–1092 (1986).
33. Kouyama, T. & Mihashi, K. *Eur. J. Biochem.* **114**, 33–38 (1981).
34. Cooper, J. A., Walker, J. B. & Pollard, T. D. *J. Muscle Res. Cell Mot.* **4**, 253–262 (1983).
35. Jahn, R., Schiebler, W. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1684–1687 (1984).
36. Benfenati, F. & Guardabasso, V. in *Principles and Methods in Receptor Binding* (eds Cattabeni, F. & Nicosia, S.) 41–63 (Plenum, New York, 1984).

A single human gene encoding multiple tyrosine hydroxylases with different predicted functional characteristics

Brigitte Grima*, Annie Lamouroux*, Claudette Boni*, Jean-François Julien*, France Javoy-Agid† & Jacques Mallet*‡

* Laboratoire de Neurobiologie Cellulaire et Moléculaire, Département de Génétique Moléculaire, Centre National de la Recherche Scientifique, F-91190 Gif-sur-Yvette, France

† Institut National de la Santé et de la Recherche Médicale, U 289, Laboratoire de Médecine Expérimentale, 91, bld de l'Hôpital, F-75013 Paris, France

Catecholaminergic systems in discrete regions of the brain are thought to be important in affective psychoses, learning and memory, reinforcement and sleep-wake cycle regulation¹. Tyrosine hydroxylase (TH) is the first enzyme in the pathway of catecholamine synthesis. Its importance is reflected in the diversity of the mechanisms that have been described which control its activity²; TH levels vary both during development and as a function of the activity of the nervous system. Recently, we deduced the complete amino-acid sequence of rat TH from a complementary DNA clone encoding a functional enzyme^{3,4}. Here we demonstrate that, in man, TH molecules are encoded by at least three distinct messenger RNAs. The expression of these mRNAs varies in different parts of the nervous system. The sequence differences observed are confined to the 5' termini of the messengers and involve alternative splicing events. This variation has clear functional consequences for each putative form of the enzyme and could represent a novel means of regulating catecholamine levels in normal and pathological neurons.

To isolate full-length cDNA clones encoding human TH, a λgt10 cDNA library was generated from a pheochromocytoma

‡ To whom correspondence should be addressed.



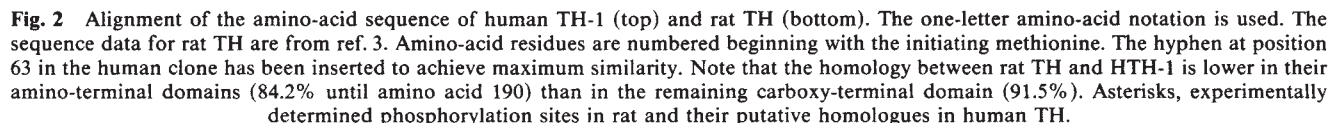


Fig. 2 Alignment of the amino-acid sequence of human TH-1 (top) and rat TH (bottom). The one-letter amino-acid notation is used. The sequence data for rat TH are from ref. 3. Amino-acid residues are numbered beginning with the initiating methionine. The hyphen at position 63 in the human clone has been inserted to achieve maximum similarity. Note that the homology between rat TH and HTH-1 is lower in their amino-terminal domains (84.2% until amino acid 190) than in the remaining carboxy-terminal domain (91.5%). Asterisks, experimentally determined phosphorylation sites in rat and their putative homologues in human TH.

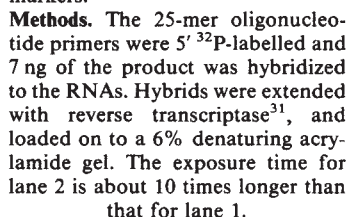


Fig. 3 Primer extension assays of 2.5 µg human pheochromocytoma poly(A)⁺ RNA with an oligonucleotide primer complementary to HTH-1 RNA: 5'-AGCCCTTGCCCTGTGGCGTGGTGGC-3' (lane 1) and an oligonucleotide primer complementary to HTH-3 RNA: 5'-TCCAGGCCACGGAGCCGTGTAGG-3' (lane 2). Lane 3, size markers.

Methods. The 25-mer oligonucleotide primers were 5' ³²P-labelled and 7 ng of the product was hybridized to the RNAs. Hybrids were extended with reverse transcriptase³¹, and loaded on to a 6% denaturing acrylamide gel. The exposure time for lane 2 is about 10 times longer than that for lane 1.

complementary to the residues underlined in line 3 of Fig. 1B yielded two faint bands of equal intensity corresponding to species of 161 and 154 bases, respectively. The second band may result from a stop in reverse transcription or from the use of a slightly different initiation site under the control of the same promoter. Because 45 bases of the extended fragments match the 5' end of the sequence of HTH-3 shown in Fig. 1B, these data suggest that HTH-3 lacks 116 bp at its 5' end to be full length. Strikingly, the first exon of HTH-1 pre-mRNA contains 119 bases and ends with an ATG codon which could conceivably correspond to that located at the 5' end of HTH-3. The latter would thus lack the entire first exon except for 3 bp. In line with this hypothesis preliminary mapping experiments of human *TH* genomic clones indicate that sequences characteristic of HTH-3 are downstream of the first exon. Alternatively, the HTH-3 pre-mRNA could result from transcription at another promoter located within an intron. Generation of protein diversity by this means has now been demonstrated in other systems^{7,8}. Analysis of additional clones of the HTH-3 type which extend further at the 5' end should allow us to discriminate between these two possibilities.

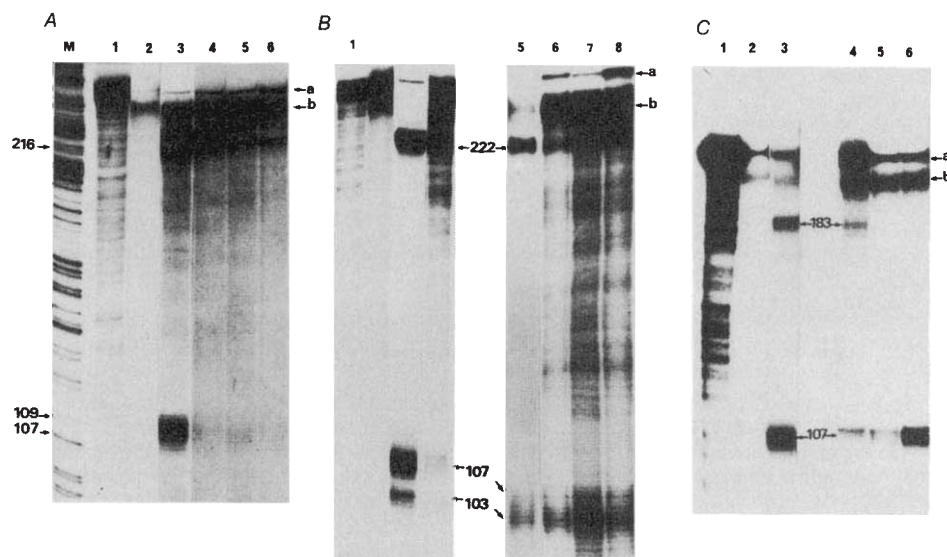
The occurrence of multiple mRNAs in the pheochromocytoma tumour was confirmed by S_1 nuclease mapping experiments using uniformly labelled single-stranded probes. The 5' ends of the three cDNAs were subcloned in M13 vectors by exploiting the cloning *Eco*RI site, the *Sma*I site located at the 5' terminus of HTH-3 and the internal *Sma*I site common to all three clones, at position 100 (Fig. 1A). The calculated lengths of the three M13 inserts were therefore 216, 222 and 183 bases for HTH-1, HTH-2 and HTH-3, respectively. As expected, the probe originating from HTH-1 yielded a protected fragment of 216 bases corresponding to HTH-1 mRNA (Fig. 4A, lane 3). A second set of protected fragments of around 109 and 107 bases were associated with HTH-2 and HTH-3 mRNAs. Similarly, the two other probes yielded the expected pattern corresponding to protected fragments of 222, 183, 107 and 103 bases (Fig. 4B and C). Analysis of RNAs purified from a human adrenal gland yielded a similar pattern of protected fragments with the three probes, indicating that the expression of the multiple *TH* mRNAs was not unique to tumour cells (Fig. 4). These experiments also revealed that type-2 mRNA dominated, both in the pheochromocytoma and in adrenal glands, the relative amounts of HTH-1, HTH-2 and HTH-3 mRNAs being ~1, 2 and 0.4, respectively.

The distribution of HTH mRNA in the dopaminergic substantia nigra and the noradrenergic locus coeruleus structures of human post mortem brains was analysed by protection experiments. HTH-1 and HTH-2 mRNAs were present in both tissues. However, in neither tissue could we detect HTH-3 mRNA. It

for the inclusion of 12 bp between nucleotides 90 and 91 (Fig. 1A and B), suggesting that the corresponding mRNA results from alternative splicing. This possibility was confirmed by comparing HTH-1 and HTH-2 sequences with that of a partial rat genomic clone encoding TH (Fig. 1B and C). The latter contains a TATA box at position -64 and three putative initiation sites for transcription at positions -33 to -35 were revealed by primer extension (Fig. 1C). The first exon extends to nucleotide 90, corresponding precisely to the point of divergence between HTH-1 and HTH-2. Typical splice acceptor sequences are found at the boundaries of the following intron⁶. Figure 1B and C shows that the 12-bp insertion is homologous to the 5' end of intron 1, suggesting that it derives from a 3' extension of exon 1.

The first exon shared by HTH-1 and HTH-2 pre-mRNAs is replaced in HTH-3 by a sequence showing no homology to the other two. This sequence begins with an AUG codon that specifies a reading frame matching the downstream coding sequences common to all three HTH mRNAs; it is not however established whether this AUG is indeed the initiation codon. Primer extension experiments using a 25-mer oligonucleotide

Fig. 4 S_1 Nuclease analysis of the 5' end of HTH mRNAs. The analysis was carried out with 5'-end probes complementary to HTH-1 (A), HTH-2 (B) and HTH-3 RNAs (C). The lengths of S_1 -protected fragments are indicated. Lane M, marker DNA fragments; lanes 1, undigested probes; lanes 2, 10 μ g yeast transfer RNA. In A and C: lanes 3, 18 μ g human pheochromocytoma total RNA; lanes 4, 2 μ g human adrenal medulla poly(A)⁺ RNA; lanes 5, 20 μ g human locus coeruleus total RNA; lanes 6, 20 μ g human substantia nigra total RNA. In B: two series of experiments (lanes 1-4 and lanes 5-8) have been performed; lanes 3 and 5, 18 μ g human pheochromocytoma total RNA; lanes 4 and 6, 2 μ g human adrenal medulla poly(A)⁺ RNA; lane 7, 20 μ g human locus coeruleus total RNA; lane 8, 20 μ g human substantia nigra total RNA. Note that in the first series of experiment the fragments corresponding to 103 and 107 bp are clearly distinguishable. Bands marked 'a' correspond to undigested probes and bands 'b' probably correspond to a partial digestion of the unhybridized probes. The exposure times were about 10 times shorter for experiments carried out with pheochromocytoma RNA than for the others.



Methods. Total RNA and poly(A)⁺ RNA was prepared as in Fig. 1. The *EcoRI*-*SmaI* fragments corresponding the 5' end of clones HTH-1 and HTH-2 and the *SmaI* fragment corresponding to the 5' part of the clone HTH-3 were subcloned into M13mp8 vectors. Each single-stranded DNA (1 μ g) was hybridized with M13 sequencing primer (P.L. Biochemicals) and the labelled complementary strands were generated with Klenow polymerase. The double-stranded DNAs were digested with *EcoRI* and the labelled single-stranded inserts were isolated on a 5% denaturing acrylamide gel³². The three purified probes corresponding to the HTH-1, HTH-2, and HTH-3 cDNAs are 264, 270 and 231 bp long, respectively, including 48 bp from the M13 linker and primer and a 107-bp common fragment. The probes (100,000 c.p.m.) were hybridized to RNA in 80% formamide at 50 °C for 12 h, and then digested with S_1 nuclease (400 U ml⁻¹) at 37 °C for 1 h. The samples were extracted, precipitated and loaded onto a 6% denaturing acrylamide gel. To determine the relative abundance of the three TH mRNAs we took into account only the signals resulting from D-loop digestion because the single cleavage site in the DNA that forms an R-loop is not easily accessible. This explains the weak signal in lanes 6-8 of A. Human adrenal glands were obtained from Dr E. Laforge (Hôpital Pellegrin, Bordeaux). Human brain specimens were obtained from the brains of two patients who died with no evidence of neurological or psychiatric disease (Hôpital Pitié-Salpêtrière, Paris). Less than 2 h after autopsy, the substantia nigra pars compacta and the locus coeruleus were dissected out under a magnifying glass and immediately frozen in liquid nitrogen. Tissue samples were stored at -70 °C until biochemical analysis.

is conceivable that this finding results from differential post-mortem stability of the three mRNAs. This seems unlikely, because different relative quantities of the multiple TH mRNAs were also observed in rat tissues that were collected immediately after killing the rat (data not shown). Whether HTH-3 is completely absent from the central nervous system and whether HTH-1 and HTH-2 are co-expressed in the same neurons await *in situ* hybridization experiments⁹.

The sequences of HTH mRNAs predict the existence of three different polypeptides. This deduced structural variation could therefore help to explain previous biochemical and kinetic studies suggesting that TH might exist in different forms in different tissues¹⁰⁻¹², although the observed heterogeneity could also result from variation in the purity of enzyme preparations, partial degradation during isolation or post-translational modifications.

Analysis of the primary sequence of the three molecular forms of TH we have discovered allows us to make certain predictions concerning their functional diversity. First, because the differences between the three forms are restricted to their amino termini, the catalytic site, which has been shown to be in the carboxy-terminal domain^{3,13}, must be conserved in all three. Secondly, phosphorylation of multiple sites in the amino-terminal region¹⁴ has been shown experimentally to modulate the activity of TH in the short term². Three of the serine residues implicated in the regulation are conserved, at least in HTH-1 and HTH-2. Strikingly, the addition of Val-Arg-Gly-Gln in HTH-2 endows the adjacent serine with the potential for calmodulin-dependent phosphorylation^{15,16}. Thirdly, HTH-3 contains a stretch of uncharged residues which might explain the observation that TH activity is associated with the membrane of secretory vesicles

in bovine chromaffin cells¹⁷. In addition to introducing these functional motifs, the sequence differences observed may of course also have purely structural consequences that affect the activity of the enzyme.

Long-term modulation of TH activity has been observed in a variety of experimental or pathological situations. In the light of our results, at least part of these changes could result from differential expression of the three forms. This could occur in the context of an overall increase in TH transcription, as has been demonstrated for the case of trans-synaptic induction occurring as a result of prolonged nerve activity¹⁸. Alternatively, in the absence of a change in protein levels, modulation of the relative abundance of the three forms could explain the result of Lewander *et al.*¹⁹ in the rat locus coeruleus. They proposed that the existence of a novel and more active TH could account for the delayed effect of receptor activation on TH activity in this system. In human disorders such as Parkinson's disease, in which surviving dopaminergic neurons apparently compensate for the loss of others by an increase in TH activity per cell²⁰, similar considerations could also be important.

The human TH gene was assigned to chromosome 11p15 in the vicinity of the Harvey-ras-1 and insulin genes, their relative positions being TH, Harvey-ras-1, insulin and the centromere in that order^{21,22}. Recently, an autosomal dominant form of manic-depressive psychosis has also been shown to be linked to the insulin and more strongly to the Harvey-ras-1 genes²³. Taken together, these results suggest that mutations in the TH gene may be important in the aetiology of this disease. A better understanding of the molecular aspects of TH expression may help to evaluate the function of catecholamines in affective disorders.

We thank N. Faucon Biguet for her help with S₁ mapping experiments, C. Henderson for advice and suggestions on the manuscript, F. Grosveld and his collaborators (NIMR, Mill Hill, London) for their hospitality and help in constructing the rat cosmid library. Sequence data treatments were performed using computer facilities at CITI 2 in Paris with the help of the French Ministère de la Recherche et de l'Enseignement Supérieur (Programme mobilisateur 'Essor des Biotechnologies'). This work was supported by grants from the Centre National de la Recherche Scientifique, the Association pour la Recherche sur le Cancer and Rhône-Poulenc Santé. B.G. received a fellowship from the Ligue Nationale Française contre le Cancer. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00270.

Received 24 December 1986; accepted 9 March 1987.

- Cooper, J. R., Bloom, F. E. & Roth, R. H. *The Biochemical Basis of Neuropharmacology* (Oxford University Press, 1982).
- Zigmond, R. E. *Trends Neurosci.* **8**, 63-69 (1985).
- Grima, B., Lamouroux, A., Blanot, F., Faucon Biguet, N. & Mallet, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 617-621 (1985).
- Horellou, P., Guilbert, B., Leviel, V. & Mallet, J. *FEBS Lett.* **205**, 6-10 (1986).
- Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459-472 (1982).
- Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. & Ogata, K. *Nature* **308**, 333-338 (1984).
- Stanley, E., Metcalf, D., Sobieszczuk, P., Gongh, N. M. & Dunn, A. R. *EMBO J.* **4**, 2569-2573 (1985).
- Berod, A., Faucon Biguet, N., Dumas, S., Block, B. & Mallet, J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Joh, T. H. & Reis, D. J. *Brain Res.* **85**, 146-151 (1975).
- Oka, K., Ashiba, G., Sugimoto, T., Matsuura, S. & Nagatsu, T. *Biochim. biophys. Acta* **706**, 188-196 (1982).
- Acheson, A. L., Kapatos, G. & Zigmond, M. J. *Life Sci.* **28**, 1407-1420 (1981).
- Mallet, J. *et al.* in *The Role of RNA and DNA in Brain Function; a Molecular Biological Approach* (eds Giuditta, A., Kaplan, B. & Zonzely-Neurath, C.) 57-70 (Nijhoff, Boston, 1986).
- Campbell, D. G., Hardie, D. G. & Vulliet, P. R. *J. biol. Chem.* **261**, 10489-10492 (1986).
- Pearson, R. B., Woodgett, J. R., Cohen, P. & Kemp, B. E. *J. biol. Chem.* **260**, 14471-14476 (1985).
- Cohen, P. *Eur. J. Biochem.* **151**, 439-448 (1985).
- Treiman, M., Weber, W. & Gratzl, M. *J. Neurochem.* **40**, 661-669 (1983).
- Faucon Biguet, N., Buda, M., Lamouroux, A., Samolyk, D. & Mallet, J. *EMBO J.* **1**, 287-291 (1986).
- Lewander, T., Joh, T. H. & Reis, D. J. *Nature* **258**, 440-441 (1975).
- Bokobza, B., Ruberg, M., Scatton, B., Javoy-Agid, F. & Agid, Y. *Eur. J. Pharmac.* **99**, 167-175 (1984).
- Craig, S. P., Buckle, V. J., Lamouroux, A., Mallet, J. & Craig, I. W. *Cytogenet. Cell Genet.* **42**, 29-32 (1986).
- Moss, P. A. H., Davies, K. E., Boni, C., Mallet, J. & Reeders, S. T. *Nucleic Acids Res.* **14**, 9927-9932 (1986).
- Egeland, J. A. *et al.* *Nature* **325**, 783-787 (1987).
- Montminy, M., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6682-6686 (1986).
- Comb, M., Birnberg, N. C., Seasholtz, A., Herbert, E. & Goodman, H. M. *Nature* **323**, 353-356 (1986).
- Lomedico, P. & Saunders, G. *Nucleic Acids Res.* **3**, 381-391 (1976).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Techniques: a Practical Approach* Vol. I (ed. Glover, D.) 49-78 (IRL Press, Oxford, 1984).
- Grosveld, F. G. *et al.* *Nucleic Acids Res.* **10**, 6715-6732 (1982).
- Grosveld, F. G., Dahl, H. H. M., De Boer, E. & Flavell, R. A. *Gene* **133**, 227-237 (1981).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- McKnight, S. L. & Kingsbury, R. *Science* **217**, 316-324 (1982).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).

An inducible transcription factor activates expression of human immunodeficiency virus in T cells

Gary Nabel* & David Baltimore

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA and *Howard Hughes Medical Institute

Human immunodeficiency virus (HIV) production from latently infected T lymphocytes can be induced with compounds that activate the cells to secrete lymphokines^{1,2}. The elements in the HIV genome which control activation are not known but expression might be regulated through a variety of DNA elements. The *cis*-acting control elements of the viral genome are enhancer and promoter regions. The virus also encodes *trans*-acting factors specified by the *tat*-III (refs 3-6) and *art* genes⁷. We have examined whether products specific to activated T cells might stimulate viral transcription by binding to regions on viral DNA. Activation of T cells, which increases HIV expression up to 50-fold, correlated with induction of a DNA binding protein indistinguishable from a recognized transcription factor, called NF- κ B (ref. 8), with binding sites in the viral enhancer. Mutation of these binding sites abolished inducibility. That NF- κ B acts in synergy with the viral *tat*-III gene product to enhance HIV expression in T cells may have implications for the pathogenesis of AIDS (acquired immune deficiency syndrome).

To determine if T-cell activation can increase viral transcription, we transfected Jurkat cells, a T4⁺ human T-leukaemia cell line, with plasmid HIV-CAT, also called pUR-III (ref. 9), which contains the HIV enhancer and promoter region upstream of the bacterial chloramphenicol acetyl transferase (*CAT*) gene (Fig. 1). Because the Jurkat line, like normal T cells, requires two signals to synthesize optimal amounts of lymphokines¹⁰, the cells were incubated with phorbol myristate acetate (PMA, 4- β -phorbol-12- β -myristate- α -acetate) and phytohemagglutinin (PHA) for different periods after transfection: all cells were incubated for 44 h but were treated for the last 0, 2.5, 5 or

20 h with the activators. In a time-dependent fashion, the activity of CAT was greatly augmented by the activators (Fig. 2a), being increased a maximum of 10- to 50-fold in various experiments.

To examine whether the increase in expression was mediated by an identifiable, inducible protein, nuclear extracts were prepared from resting or activated Jurkat cells, DNA probes were derived from the enhancer region of HIV^{9,11}, and the interaction between them was studied using the electrophoretic mobility shift assay¹²⁻¹⁵. One probe from the U3 region of the viral long terminal repeat (see Fig. 1) bound to factors in nuclear extracts from activated Jurkat cells, but not from unstimulated cells (Fig. 3a, lanes 1 and 2).

The probe that bound to the factor came from a region previously shown by deletion mapping^{9,11} to contain the HIV enhancer. The enhancer region contains two copies of an 11-base pair (bp) sequence, κ B, also found in the κ -immunoglobulin gene enhancer region. In B cells, the transcription factor NF- κ B binds to this sequence and appears to be responsible for the induction of κ -gene transcription by lipopolysaccharide or phorbol esters¹⁶. This site in the HIV enhancer region is found once exactly as found in the κ gene enhancer and once with a single base-pair alteration (see Fig. 1). When the HIV enhancer fragment was mutated at bases critical for binding to NF- κ B in B cells (ref. 8; see Fig. 1), nuclear extracts from activated Jurkat cells no longer showed binding activity (Fig. 3a, lane 3). The site-specificity of this inducible binding factor was further confirmed by competition studies. When an excess of an unlabelled DNA fragment containing the κ -light chain enhancer was present in the binding reaction, no complex with the inducible factor was formed on the HIV enhancer probe (Fig. 3b, lanes 2 and 3). Inclusion of a DNA fragment from the interleukin-2 (IL-2) promoter, lacking the κ B sequence, did not inhibit formation of this complex (lanes 4 and 5), further confirming specificity of the complex for the 11-bp κ B motif. The second major band did not change upon induction, was not competed away by the unlabelled cognate fragment, and showed no protection in methylation interference assays (data not shown), indicating non-specific binding.

To study whether binding to this enhancer sequence increased viral transcription, we transfected Jurkat cells with normal HIV-CAT or HIV-CAT containing mutations at both κ B enhancer sites. Although normal HIV-CAT transfections showed a nearly

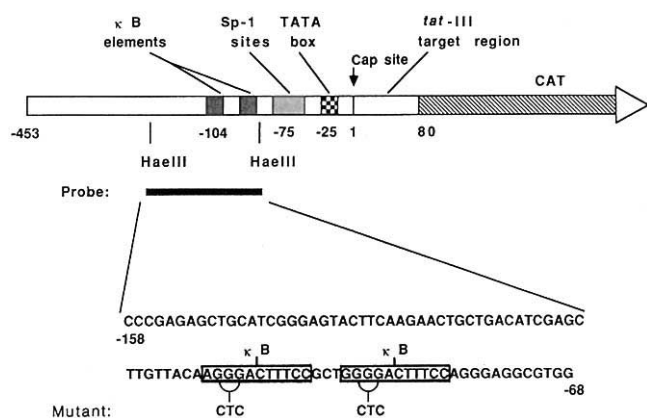


Fig. 1 Location of inducible enhancer element in HIV-LTR and sequence of normal and mutant inactive plasmids. Relevant sites within the LTR are identified. Altered bases in the mutant plasmid are shown below the wild-type sequence.

Methods. Sequence and position of the HIV enhancer have been obtained previously by deletion mapping^{9,11}. Mutant plasmid containing three base-pair changes in critical contact residues⁸ was constructed using a modification of the oligonucleotide-directed gap heteroduplex technique. HIV-CAT (1 μ g) was digested with *Bam*-HI or *Bgl*II, purified on a 0.5% low-melt agarose gel, phenol extracted, chloroform treated, ethanol precipitated, washed with 70% ethanol and resuspended in 10 μ l of 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Linear *Bam*-HI digested fragment (8 μ l) was incubated with 'gapped' *Bgl*II-treated fragment (4 μ l) in 0.2 M KOH, 0.02 M EDTA in a final volume of 30 μ l for 10 min at 37 °C. The volume was expanded to 120 μ l, with a final concentration of 0.1 M Tris-HCl pH 7.2, 50% formamide, and incubated at room temperature overnight. The reaction mixture was precipitated with ethanol in 0.3 M sodium acetate and washed with 70% ethanol. The precipitate was resuspended in 20 μ l, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 M NaCl, and 100 ng of a synthetic oligonucleotide identical to positions -119 to -77 of HIV enhancer, except for the modification of six base pairs (Fig. 1), heated to 68 °C for 5 min, and incubated for 1.5 h at room temperature. The volume was expanded to 50 μ l in DNA ligase buffer²⁸, 3 μ l of DNA polymerase (Klenow fragment) and 2 μ l of T4 ligase²⁸, before transformation of bacteria. Mutant colonies were identified using the kinase-labelled oligonucleotide by the method of Grunstein and Hogness²⁹, washing filters at 68 °C in 2 $\times$ SSC. Plasmid minipreps were prepared, digested with *Bgl*II, run on 1% agarose gels, transferred to nitrocellulose³⁰ and reprobed to confirm identity. Finally, the sequence of the mutant plasmids was determined by the Maxam-Gilbert technique³¹.

Fig. 2 CAT activity directed by HIV-CAT is induced in activated Jurkat cells and induction depends on κ B sequences in the viral LTR. *a*, Jurkat cells were transfected with the HIV-CAT plasmid using DEAE-dextran³² and maintained for 44 h. PHA-P (Pharmacia) was added (2 μ g ml⁻¹) with 10 nM PMA (Sigma) at the indicated times before harvest (lanes 1-4). CAT activity was measured by determining the amount of acetylated chloramphenicol (AC) produced from ¹⁴C chloramphenicol (CM). *b*, Jurkat cells were transfected as above with the HIV-CAT plasmid (lanes 1 and 2) or a mutant containing the three base-pair changes in both κ B sequences shown in Fig. 1 (lanes 3 and 4). Uninduced cells (lanes 1 and 3) remained in culture for 44 h. Induced cells (lanes 2 and 4) were treated with PHA (2 μ g ml⁻¹) and PMA (10 nM) for 20 h before preparation of extract for CAT assay 44 h after transfection. CM and AC as in *a*.

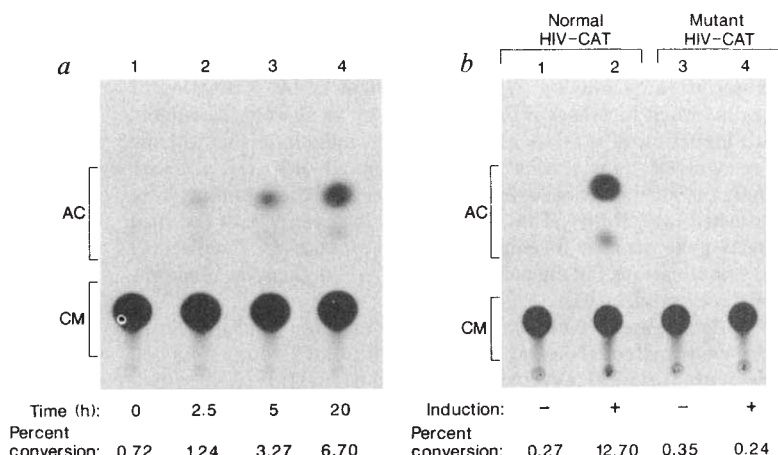
Methods. Cells (10⁷) were transfected with 10 μ g of plasmid using DEAE-dextran³² and maintained in RPMI 1640 containing 10% fetal calf serum (FCS) and penicillin-streptomycin for 44 h. Cell extracts were prepared for CAT assay and CAT activity was determined according to standard methods³³. To normalize between groups, co-transfection was carried out with a second independent indicator plasmid, pCH110, which expresses β -galactosidase activity³⁴. Protein concentration in the nuclear extract was subsequently used for standardization when it was found to correlate closely with β -galactosidase activity. Degree of conversion was determined by cutting out spots containing unreacted ¹⁴C-chloramphenicol and acetylated forms, and quantitating the amount of radioactivity in a liquid scintillation counter.

50-fold increase in CAT activity after activation (Fig. 2*b*, lanes 1 and 2), no increase above background was observed in cells transfected with mutant plasmid (lanes 3 and 4). These data suggest that binding to the κ B sequences mediates the increase in expression, and this binding may initiate viral transcription in latently infected cells that have been activated.

The *tat*-III-encoded protein of HIV is a potent activator of viral gene expression^{3-6,9}. To determine whether the cellular transcription factor NF- κ B would increase transcription in cells constitutively expressing the viral *tat*-III gene, we analysed a Jurkat cell line infected with a retrovirus vector expressing the *tat*-III gene^{17,18}. Transfection with the HIV-CAT plasmid, which contains the target region for the *tat*-III gene product (Fig. 1), led to a greater than 30-fold increase in CAT activity (Fig. 4, lanes 1 and 3). This increase was not likely to be due to differences in transfection efficiency because transfection with a plasmid containing the simian virus 40 (SV40) enhancer and promoter linked to CAT showed 4.8% and 5.1% conversion in Jurkat, and Jurkat expressing *tat*-III, respectively. When these cells were activated following transfection, a large further increase in CAT activity was observed (lane 2), suggesting that the *tat*-III gene product can act in concert with the NF- κ B transcription factor. The induction level in Fig. 4, lane 2, is underestimated because the assay was out of its linear range: when measured in a linear range, a nearly 20-fold stimulation was found. Thus, NF- κ B can exert its full effect in the presence of *tat*-III and both gene products have independent effects on CAT accumulation in activated HIV-infected cells.

We have shown that the activation of a T-cell line by mitogen leads to the expression of an inducible DNA binding protein, NF- κ B. This DNA binding protein is found normally in cells of the B lineage which express immunoglobulin light chain⁸. It is also found in a variety of cells following activation with phorbol esters¹⁶, although the target gene(s) affected by this factor in non-B cells remain unknown. Similarly, which cellular gene(s) are normally induced by NF- κ B in activated T cells is unknown. The κ B enhancer motif is not found in regulatory regions of interleukin-2, interleukin-3 or γ -interferon. The possibility remains that it may activate yet another inducible lymphokine. This sequence has been found in several viruses which infect primate cells, including SV-40 and cytomegalovirus^{19,20}, and cellular activation may enhance expression of SV40 in some cases as well²¹. Furthermore, herpes simplex virus and other DNA viruses have been shown to activate HIV transcription, raising the possibility that viruses may activate NF- κ B^{22,23}.

In an HIV-infected T cell, NF- κ B may serve two functions.



To normalize between groups, co-transfection was carried out with a second independent indicator plasmid, pCH110, which expresses β -galactosidase activity³⁴. Protein concentration in the nuclear extract was subsequently used for standardization when it was found to correlate closely with β -galactosidase activity. Degree of conversion was determined by cutting out spots containing unreacted ¹⁴C-chloramphenicol and acetylated forms, and quantitating the amount of radioactivity in a liquid scintillation counter.

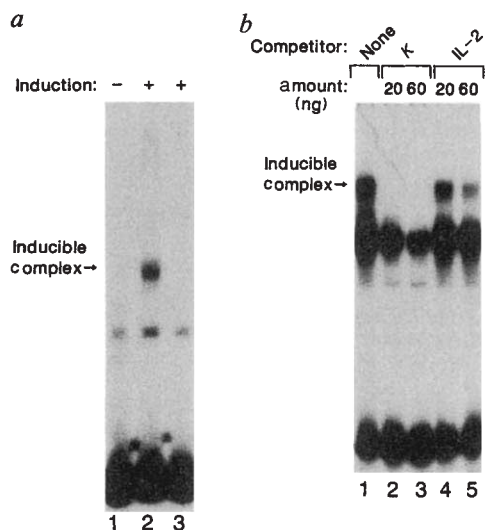


Fig. 3 Induction of enhancer binding factor in Jurkat cells. *a*, Analysis of the κ B binding factor using an electrophoretic mobility shift assay. Nuclear extracts from Jurkat cells, either uninduced (-) or activated with PMA and PHA (+), were incubated with a radiolabelled DNA probe containing either the normal HIV enhancer (lanes 1 and 2) or a mutant HIV enhancer (lane 3) with three base-pair alterations in each κ B sequence (see Fig. 1). *b*, Analysis of binding-site specificity of the induced factor by competition with known DNA fragments. Nuclear extracts from induced Jurkat cells were incubated with a radiolabelled DNA probe from the HIV enhancer alone (lane 1) or in the presence of indicated amounts of the unlabelled fragment from the κ enhancer (κ) (lanes 2 and 3) or a control fragment of the same size from the interleukin-2 promoter (IL-2) (lanes 4 and 5).

Methods. Nuclear extracts were prepared according to the method of Dignam *et al.*³⁵. Activated Jurkat cells were incubated in RPMI 1640 containing 10% FCS for 2 h before preparation of nuclear extract. The procedure for electrophoretic mobility shift assay has been described¹⁵: here the gel running buffer was 50 mM Tris base, 380 mM glycine, and 2 mM EDTA at pH 8.5. From 5 to 10 μ g of nuclear protein was used in DNA binding reactions. Radiolabelled probe was prepared from normal or mutant fragment (Fig. 1). Unlabelled competitor fragment (κ) was prepared from an *Alu*I fragment containing the entire κ -light chain enhancer (475 bp) subcloned into the *Sma*I site of pUC 13 and removed at flanking sites in the polylinker⁸. An *Rsa*I fragment of IL-2 promoter (IL-2) was subcloned into pUC13 and prepared similarly.

First, it may activate virus from its dormant state, binding to the viral enhancer region and increasing transcription of the viral genome. Second, after viral transcription leads to the synthesis of more *tat*-III protein, NF- κ B could act in concert with it to increase virus expression further. The mechanism by which the *tat*-III protein activates virus expression is controversial. Although some reports suggest a post-transcriptional mechanism^{7,11,24,25} to explain transactivation, others find an effect on transcription as well^{3,9,24,26}. Our data shows that NF- κ B independently enhances expression in the presence of *tat*-III. Previous studies have demonstrated the presence of another transcription factor binding site, SP-1, in the HIV promoter²⁷. Although this sequence is necessary for viral transcription, it does not appear to be solely responsible for activation.

These data have implications for the pathogenesis of AIDS. It is likely that activation of T cells by infectious agents induces NF- κ B, in a fashion similar to its induction by PMA and PHA, and that this transcription factor acts to increase virus production. Invasion by viral, bacterial and protozoan organisms in patients would therefore pose a risk beyond the immediate

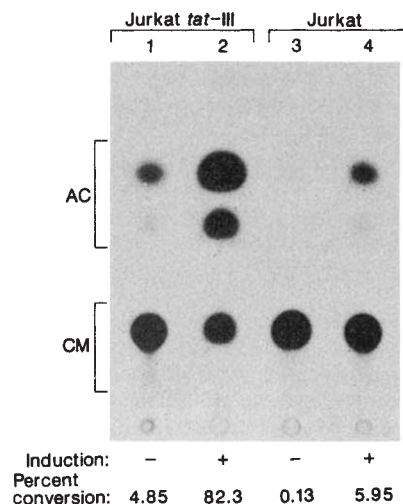


Fig. 4 HIV-CAT activity is induced in activated Jurkat cells stably expressing the *tat*-III gene. Samples of 10^7 Jurkat cells stably expressing the *tat*-III gene (lanes 1 and 2) or normal Jurkat cells (lanes 3 and 4) were transfected with 10 μ g HIV-CAT plasmid using DEAE-dextran³² and maintained in culture (lanes 1 and 3) or incubated for 20 h with PHA and PMA (lanes 2 and 4) before being collected at 44 h (Fig. 2 legend). Conversion of ¹⁴C-chloramphenicol to the acetylated forms was quantitated as in Fig. 2.

infection. NF- κ B induction by such immune stimulation might activate virus in dormant infected T cells, accelerating its spread in the early stages of the disease.

We thank Drs C. Rosen, J. Sodroski and W. Haseltine for providing HIV plasmids and the Jurkat *tat*-III cell line; Dr R. Sen for provision of κ -light chain DNA fragments, Drs M. Muesing, A. Bernards and G. Daley for advice and helpful comments on the manuscript, and G. Pierce for help in preparation of the manuscript.

Received 5 December 1986; accepted 20 February 1987.

1. Zagury, D. *et al.* *Science* **231**, 850-853 (1986).
2. Harada, S. *et al.* *Virology* **154**, 249-258 (1986).
3. Sodroski, J., Rosen, C. & Haseltine, W. *Science* **227**, 171-173 (1985).
4. Arya, S., Guo, C., Josephs, S. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
5. Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. *Science* **229**, 74-77 (1985).
6. Sodroski, J. *et al.* *Nature* **321**, 412-417 (1986).
7. Rosen, C. A. *et al.* *Nature* **319**, 555-559 (1986).
8. Sen, R. & Baltimore, D. *Cell* **46**, 705-716 (1986).
9. Rosen, C. A., Sodroski, J. & Haseltine, W. A. *Cell* **41**, 813-823 (1985).
10. Weiss, A. *et al.* *J. Immun.* **133**, 1-7 (1984).
11. Muesing, M., Smith, D. H. & Capon, D. J. *Cell* **48**, 691-701 (1987).
12. Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
13. Garner, M. M. & Rezin, A. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
14. Strauss, F. & Varshavsky, A. *Cell* **37**, 889-901 (1984).
15. Singh, H., Sen, R., Baltimore, D. & Sharp, P. *Nature* **319**, 154-158 (1986).
16. Sen, R. & Baltimore, D. *Cell* **47**, 921-928 (1986).
17. Dayton, A. *et al.* *Cell* **44**, 941-947 (1986).
18. Greene, W. C. *et al.* *Science* **232**, 877-880 (1986).
19. Sassone-Corsi, P., Wildeman, A. & Chambon, P. *Nature* **313**, 458-463 (1985).
20. Boshart, M. *et al.* *Cell* **41**, 521-530 (1985).
21. Imbra, R. & Karin, M. *Nature* **323**, 555-557 (1986).
22. Mosca, J. D. *et al.* *Nature* **325**, 67-70 (1987).
23. Gendelman, H. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 9759-9763 (1986).
24. Cullen, B. R. *Cell* **46**, 973-982 (1986).
25. Feinberg, M. *et al.* *Cell* **46**, 807-817 (1986).
26. Okamoto, T. & Wong-Staal, F. *Cell* **47**, 29-35 (1986).
27. Jones, K. A., Kodonaga, J. T., Luciw, P. A. & Tjian, R. *Science* **232**, 755-759 (1986).
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
29. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
30. Smith, G. & Summers, M. D. *Analyt. Biochem.* **109**, 123-129 (1980).
31. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
32. Queen, C. & Baltimore, D. *Cell* **33**, 741-748 (1983).
33. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
34. Hall, C., Jacob, P. E., Ringold, G. & Lee, F. J. *Molec. appl. Genet.* **2**, 101-109 (1983).
35. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).

Direct activation of resting T lymphocytes by human T-lymphotropic virus type I

Louis Gazzolo & Madeleine Duc Dodon

Laboratoire d'Immuno-Virologie Moléculaire, UM 380030 du CNRS, Faculté de Médecine Alexis Carrel, 69372 Lyon Cedex 08, France

The mitogenic or antigenic stimulation of T lymphocytes in the presence of accessory cells results in the synthesis of interleukin-2 (IL-2), which, on interacting with *de novo* synthesized receptors on the membrane, induces the proliferation and differentiation of T cells¹⁻⁵. T lymphocytes are the preferential target cells of human T-lymphotropic virus type I (HTLV-I). This virus was isolated from leukaemic cells of patients with adult T-cell leukaemia⁶⁻⁸. No viral transcription was detected in these leukaemic cells, indicating that consistent expression of HTLV-I is not needed for maintenance of the neoplastic state⁹. After a short time in culture, these leukaemic cells produce viral particles which immortalize normal T cells. Hence, HTLV-I might be an important agent not only in the development, but also in the initiation, of this lymphoproliferative disease. This possibility was sustained by our previous observations indicating that normal T lymphocytes incubated with HTLV-I are able to form colonies in soft agar, seven days later, in the absence of exogenous IL-2¹⁰. These results led us to explore further the immediate effects of viral infection on purified resting T lymphocytes. Here, we present evidence that HTLV-I is able to activate T lymphocytes. Indeed, binding of viral particles to these cells induces IL-2 production and IL-2 receptor expression, and triggers T-cell proliferation. This initial activation, which appears to be independent of accessory cells, may be relevant in understanding the role of HTLV-I in the aetiology of adult T-cell leukaemia.

To assess the proliferative response of T lymphocytes to HTLV-I, we added virus produced by a T-lymphoblastoid cell line (C91/PL)¹¹ to cells at three different steps of preparation: to mononuclear cells (MNC) obtained after Ficoll-Hypaque centrifugation of peripheral blood, to T cells obtained after removal of adherent and phagocytic cells from MNC and subsequent E-rosetting, and finally to DR-depleted T cells obtained after elimination of HLA-DR-bearing cells. Proliferation was assayed by the level of DNA synthesis on day 5. HTLV-I induced cell proliferation in all three types of culture (Fig. 1a). The stimulation index (SI, the ratio of thymidine incorporation by HTLV-I-infected cells to that by uninfected cells) was found to be 61 for MNC, 127 for T cells and 160 for the DR-depleted population.

We then compared the effect of HTLV-I to activation by phytohaemagglutinin (PHA), a mitogenic lectin that requires the presence of accessory cells to activate T cells^{4,5}. PHA induced a strong proliferative response only in MNC ($SI_{PHA} = 251$). For T cells, proliferation was decreased strongly, correlated with the progressive removal of accessory cells. Thus, the SI_{PHA} of 168 is comparable to that observed with HTLV-I ($SI_{HTLV} = 127$). As expected, DR-depleted T cells were quite unresponsive ($SI_{PHA} = 2$), in striking contrast with the result obtained after incubation with HTLV-I ($SI_{HTLV} = 160$).

To determine whether this proliferative response was specific to HTLV-I, we infected cells of each type with HIV (human immunodeficiency virus), another human retrovirus (associated with AIDS), obtained from a T-lymphoblastoid cell line permanently infected with the virus¹². No proliferation was observed in any type of culture after addition of this virus.

The mitogenic activity we observed was not caused by mycoplasma contamination, as the cell line used was consistently negative in tests for mycoplasma. Moreover, the mitogenic effect was not mediated by soluble factors present in the growth

Table 1 Proliferation of T lymphocytes in presence of HTLV-I

Culture	Experiment 1	Experiment 2
T-cells	100 ± 30	110 ± 20
+HTLV-I(C91/PL)	12,300 ± 1,500	73,000 ± 8,800
+HTLV-I(HUT 102)	4,100 ± 600	ND
+Control(C8166/45)	60 ± 30	580 ± 130
+PHA	16,500 ± 5,500	36,300 ± 2,300
DR-depleted	120 ± 40	80 ± 20
+HTLV-I(C91/PL)	19,400 ± 6,200	86,200 ± 19,700
+HTLV-I(HUT 102)	7,000 ± 500	ND
+HTLV-I(MT2)	8,900 ± 2,300	ND
+Control(C8166/45)	60 ± 10	840 ± 620
+PHA	260 ± 7	6,800 ± 1,760
CD4 ⁺ -DR-depleted	ND	430 ± 210
+HTLV-I(C91/PL)	ND	50,700 ± 4,300
+PHA	ND	2,100 ± 590
CD8 ⁺ -DR-depleted	ND	770 ± 260
+HTLV-I(C9/PL)	ND	51,500 ± 16,400
+PHA	ND	1,200 ± 200

T lymphocytes and DR-depleted cells were prepared as indicated in Fig. 1. CD4⁺ and CD8⁺ T cells were isolated from DR-depleted T cells in a two-step cytotoxic assay by incubating them first with monoclonal antibody OKT8 and with OKT4 (0.4 per µg 10⁶ cells; Becton Dickinson) respectively, and then in new-born rabbit complement. HTLV-I particles were prepared from culture media of three HTLV-I-producing cell lines (C91/PL, HUT102, MT2) as Fig. 1. Pelleted material obtained after ultracentrifugation of medium of an HTLV-transformed, but non-virus producing, cell line (C81-66/45) was used as a control. PHA was added at a final concentration of 1 µg ml⁻¹. Incorporation of ³H-thymidine was measured on day 5 of the culture during the last 6 hours. Incorporation by triplicate cultures is reported as mean ± s.d. ND, not determined.

medium because the supernatant obtained after pelleting viral particles was unable to initiate T-cell proliferation. Finally, the serum obtained from one HTLV-I-positive healthy person (possessing antibodies against viral *gag* and *env* proteins as determined by electroblot assays) was able to neutralize the proliferative activity of HTLV-I particles in a dilution-dependent fashion (Fig. 1b). Moreover, we have recently observed that a monoclonal antibody, designated 0.5α (ref. 13), recognizing epitopes of the viral envelope glycoprotein gp46, was able to neutralize the proliferative ability of these viral particles. The addition of 5 µg of the 0.5α antibody led to a decrease in the thymidine incorporation by 50%, confirming that viral particles are solely responsible for initiating T-cell activation. In addition, viral particles obtained from other HTLV-I-producing cell lines (HUT102, MT2)^{14,15} were able to induce cell proliferation in T-cell or DR-depleted cultures (Table 1). Among the DR-depleted T cells, CD4⁺ T cells and CD8⁺ T cells were equally responsive to the proliferative potential of HTLV-I particles. Conversely, when T cells were incubated with the pelleted material obtained from an HTLV-I-transformed T-cell line that did not produce virus (C81-66/45)¹⁶, no proliferative activity could be measured.

These experiments indicate that T-cell proliferation induced by HTLV-I particles did not require accessory cells. To test further for possible involvement of accessory cells, DR-depleted T cells were incubated in presence of HTLV-I at a density of 10⁵–10³ cells per well, with or without 5 × 10³ non-T accessory cells per well. This number of accessory cells was able to induce proliferation of purified T cells stimulated with PHA. The T-cell response was essentially density independent and the addition of accessory cells did not significantly increase the proliferative response to HTLV-I (Table 2). However, that only a small amount of proliferation was seen with 10⁴ T cells per well in the absence of accessory cells suggests that the frequency of responding T cells is a factor limiting the proliferative effect of HTLV-I. In fact, at the same cell concentration, the addition of non-T accessory cells did restore the PHA response of DR-depleted T cells.

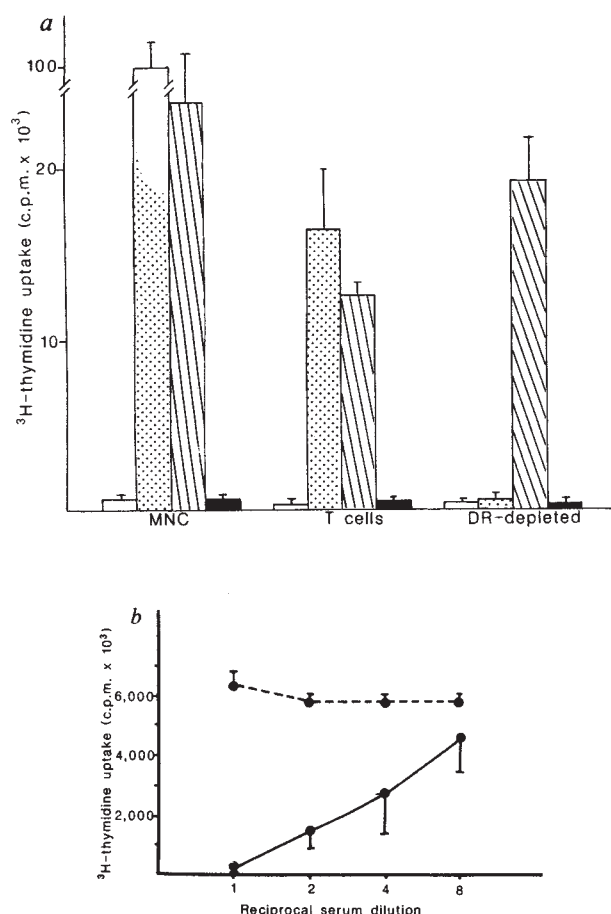


Fig. 1 *a*, Proliferative response of mononuclear cells (MNC), T cells and DR-depleted T cells either uninfected (open bars) or infected with HTLV-I (hatched bars) or with HIV (solid bars) or stimulated with PHA (stippled bars). *b*, Proliferative response to HTLV-I of DR-depleted T cells incubated either with HTLV-I antibody-containing serum (solid line) or with control negative serum (broken line).

Methods. In *a*, T lymphocytes were isolated from mononuclear cells (MNC) obtained after Ficoll-Hypaque centrifugation of peripheral blood from normal healthy donors (Centre de Transfusion Sanguine-Lyon). MNC were then incubated for 1 h at 37 °C in RPMI with 10% fetal calf serum (FCS) in culture flasks (Corning). The non-adherent cells were then depleted of phagocytic cells by carbonyl iron ingestion at 37 °C for 1 h, and passage over a magnet. The non-adherent non-phagocytic cells were incubated sheep red blood cells (SRBC), treated with 2-aminoethyl isothiuronium bromide (Sigma); the rosetting cells were separated after Ficoll gradient centrifugation. SRBC were removed by lysis with 0.15 M NH₄Cl buffer. The T-cell preparation was found to contain about 1% non-specific esterase-positive cells. Finally to remove residual HLA-DR positive cells, T cells were incubated with the murine monoclonal antibody lot 2b anti-HLA class II (Immunotech) at a final concentration of 0.5 µg per 10⁶ cells. After incubation for 1 h at +4 °C, the cells were centrifuged and resuspended in new-born rabbit complement (Centre de transfusion Sanguine, Besançon) at the appropriate dilution. This mixture was then incubated at 37 °C for 1 h with intermittent agitation. The DR-depleted cells were then recovered by centrifugation, washed and finally resuspended in Yssel medium²⁸ supplemented with 1% pooled human AB serum (heat-inactivated). Cells were seeded in triplicate 96-well flat-bottomed microculture plates at a ratio of 10⁵ cells per well in 50 µl of medium. HTLV-I was prepared from the culture medium of HTLV-I-producing cells (C91/PL)¹¹. HIV was prepared from the culture medium of H9/HTLV-III cell line¹². After culture for 48 h, media were collected and centrifuged at low speed to remove cells. The virus-containing supernatant was concentrated 20-fold by centrifuging for 3 h at 32,000g in a Beckman type 35 rotor, and by resuspending the viral pellet in an appropriate volume of Yssel medium. Viral preparations were added at 40 µl to each well. After overnight incubation, medium was added up to a final volume of 150 µl. PHA-P (Wellcome) was added at a final concentration of 1 µg ml⁻¹. In *b*, virus was incubated with serum for 1 h at room temperature before being added to T cells. Proliferation was assessed 5 days after the beginning of the culture. Thymidine incorporation (1 µCi per well; specific activity 5 Ci mmol⁻¹; CEA, France) during the last 6 h of culture, was measured by standard liquid scintillation counting techniques after collecting cells with a Skatron harvester. Data are reported as mean of triplicate cultures ±s.d.

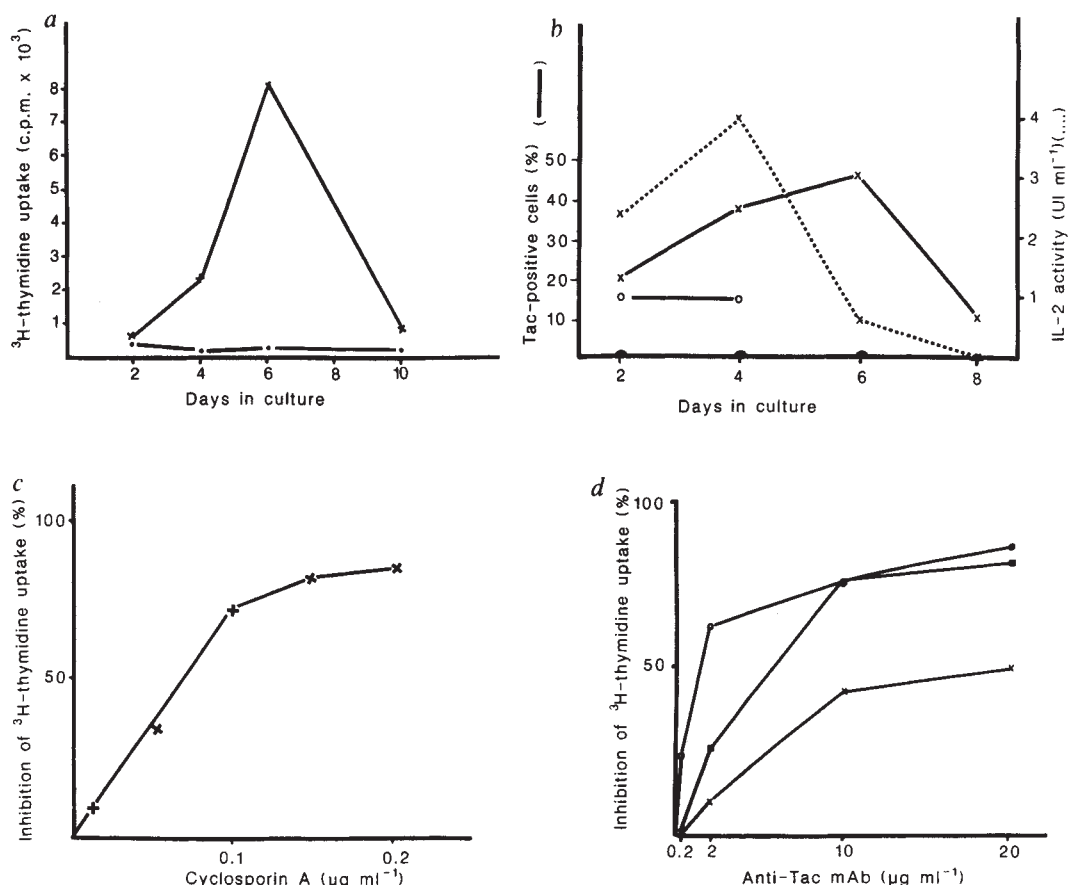
We next characterized the response of T cells to HTLV-I particles. Kinetics experiments (Fig. 2*a*) clearly showed that an increase in DNA synthesis could be observed as early as 2 days after the initiation of the cultures. This synthesis peaked on day 6; afterwards, cell proliferation decreased to values similar to those observed in control uninfected cultures. This rapid and transient proliferative response of T cells to HTLV-I might indicate that this early event could be independent of viral integration and viral replication in these cells. Therefore, we studied the effect on T-cell proliferation of HTLV-I inactivated by exposure to ultraviolet (UV) light for 3 min at a dose rate of 40 erg mm⁻² s⁻¹. It was found that UV-inactivation of HTLV-I not only did not impair the proliferative potential of the virus but increased it. Thus, in one typical experiment, the SI was 223 for cultures incubated with UV-inactivated HTLV-I compared to 160 for those incubated with the live virus. A previous investigator had noticed the mitogenic effect of live or UV-inactivated HTLV-I¹⁷ in a study performed with MNC which was limited to the three days after addition of the virus and which did not explore further this preliminary observation.

The kinetic profile, similar to that observed with antigenic T-cell stimulation, suggested that this proliferative response could be a response to IL-2 synthesis and the expression of IL-2 receptors on the T-cell membrane. We tested whether DR-

depleted cells incubated with HTLV-I could produce IL-2 and express IL-2 receptors. HTLV-I infection induced IL-2-receptor expression on 20% of lymphocytes as soon as 2 days after addition of the virus. On day 6, 47% of cells were expressing IL-2 receptors (Fig. 2*b*). This percentage decreased to 10% on day 8. Concomitantly, IL-2 activity was detected in the supernatant fluid of these cultures only at 2 and 4 days after addition of the virus. The failure to detect IL-2 on days 6 and 8 could be explained by a decrease in IL-2 synthesis and/or by binding of IL-2 to its receptors. The first possibility is more likely, because the decrease in IL-2 synthesis might correlate with the decrease in T-cell proliferation observed after 6 days of incubation. In the PHA-treated cells, only 15% of cells were expressing IL-2 receptors at 2 and 4 days after treatment. As expected, no IL-2 activity could be measured in DR-depleted cultures treated with PHA.

Proliferation of T cells in response to HTLV-I paralleled the expression of IL-2 receptors and the synthesis of IL-2. To confirm these results, inhibition experiments were performed with increasing concentrations of either monoclonal antibody against IL-2 receptor¹⁸ (anti-Tac) or cyclosporin-A¹⁹, an inhibitor known to suppress the synthesis of IL-2 and other lymphokines at the messenger RNA level, without affecting expression of IL-2 receptor. When treated with the drug

Fig. 2 *a*, Kinetics of DNA synthesis by T lymphocytes, either uninfected (●) or infected with HTLV-I (x). Medium was changed on day 6. At appropriate times, proliferation was measured by ^3H -thymidine incorporation. *b*, Kinetics of IL-2 receptor expression (solid lines; ○, in presence of PHA, X, in presence of HTLV-I on DR-depleted lymphocytes, and of IL-2 production (broken line) after addition of HTLV-I. *c*, *d*, Inhibition of proliferation of DR-depleted T cells in presence of HTLV-I after addition of either cyclosporin A (*c*) or murine monoclonal anti-Tac antibody (*d*). In *d*, HTLV-I was undiluted (x), diluted twofold (■), diluted fourfold (○).



Methods. In *a* and *b*, at indicated times, cells were collected and washed twice. Then, IL-2 receptors were detected by indirect immunofluorescence, by incubating the cells with appropriate dilutions of the murine monoclonal anti-Tac antibody for 30 min at +4 °C, washing them three times and then incubating them with fluorescein isothiocyanate-conjugated goat anti-mouse immunoglobulin (Cappel Laboratories). The percentage of anti-Tac positive cells was determined using a fluorescence microscope: 200 cells were counted per field. IL-2 in the medium of each culture was assayed by its ability to stimulate the proliferation of a murine IL-2-dependent cell line (CTLL-2)²⁹. Briefly, responder cells were cultivated in RPMI 1640, 10% FCS, 25 mm HEPES, 5×10^{-5} M 2-mercaptoethanol, 100 U penicillin, 25 $\mu\text{g ml}^{-1}$ streptomycin and 2 mM glutamine. Serial twofold dilutions of the test sample were added to 50 μl per cell of CTLL-2 cell suspension (4×10^4 cells ml^{-1}) in flat-bottomed plates. After 30 h incubation, cells were pulsed with 1 μCi ^3H -thymidine for an additional 12 h, were collected onto glass fibre filters, and measured for radioactive incorporation. A dose-response curve obtained with serial dilutions of recombinant IL-2 (Boehringer Mannheim) was used to quantify the IL-2 in these cultures. Results are the mean of triplicate cultures of one representative experiment. In *c* and *d*, DR-depleted T cells were incubated simultaneously with HTLV-I and with either cyclosporin A (Sandoz) or anti-Tac antibody (provided by Dr Waldmann), at the indicated final concentrations. In experiments performed with anti-Tac antibody, HTLV-I was used either undiluted, or diluted twofold or fourfold. Incorporation of ^3H -thymidine was determined 5 days later. In cultures incubated with the virus only, incorporation was found to be 17,000 c.p.m. for undiluted HTLV-I 9,500 c.p.m. for HTLV-I diluted twofold and 4,000 for HTLV-I diluted fourfold. Inhibition is calculated as $1[(\text{c.p.m. in treated infected cultures})/(\text{c.p.m. in untreated infected cultures})]$ expressed as a percentage. Results are the mean of triplicate cultures.

(Fig. 2c), the inhibition of proliferation was almost complete (85% with 0.2 $\mu\text{g ml}^{-1}$). When T cells were incubated with undiluted HTLV-I, the treatment of these cultures with the highest concentration of anti-Tac antibody (20 $\mu\text{g ml}^{-1}$) led to a partial inhibition of 52% (Fig. 2d). However, when T cells were incubated with dilutions of HTLV-I, the inhibition by the same concentration of anti-Tac monoclonal antibody increased by more than 80%. These results are not surprising because inhibitory effects of anti-Tac antibody on lymphocytes stimulated by PHA or pokeweed mitogen were observed only at submaximal mitogenic doses of these lectins²⁰.

We concluded first, that HTLV-I can induce the early proliferation of quiescent T cells, through the expression of IL-2 receptors and to the synthesis of IL-2; and second, that this proliferation occurs in the absence of accessory cells. Perhaps HTLV-I particles themselves supply an accessory signal. If so HTLV-I is a peculiar activator agent, and, to date, is the only virus known to induce the growth of unprimed T cells.

That virus inactivated by UV light still triggers T-cell proliferation implies that the activating event occurs before viral integration, and might be a consequence of binding of viral particles to specific receptors on the lymphocyte membrane. Several

receptors, such as the CD3/Ti complex receptor and the CD2 sheep erythrocyte receptor, are known to function in the T-cell activation process^{1,2}. It will be important to determine whether HTLV-I triggers stimulation by interacting with one of these receptors. The possibility remains, however, that the receptor involved in the proliferation process is different from that mediating the binding and the penetration of viral particles.

Most disorders related to HTLV-I are T-cell malignancies^{6,21}, although HTLV-I can infect and replicate *in vitro* in a wide variety of cells (including, fibroblasts, B cells, myeloid and endothelial cells)^{6,7}. This discrepancy might be explained by the ability of HTLV-I to trigger proliferation of quiescent T cells. Indeed, preliminary results indicate that B cells were not induced to proliferate by direct viral contact.

It remains to define the importance of this early proliferation in the initiation of the leukaemic process, and other HTLV-I-related diseases. It could be linked at least to preleukaemic stages such as the T lymphocytosis observed in pre-adult T-cell leukaemia patients²² or the presence of abnormal T lymphocytes in healthy HTLV-I carriers²³. The activation of T cells may favour the appearance of disorders in immunological processes controlled by T cells and mediators produced by these cells.

Table 2 Effect of accessory cells on T-cell proliferation induced by HTLV-I

No. of DR-depleted T-cells	Addition of accessory cells	³ H-thymidine incorporation (c.p.m.) in presence of:		
		—	HTLV-I	PHA
10 ⁵	—	84 ± 25	9,585 ± 224	167 ± 61
	+	431 ± 125	10,847 ± 385	12,720 ± 1,362
5 × 10 ⁴	—	81 ± 22	3,150 ± 632	95 ± 3
	+	155 ± 37	4,700 ± 1,325	13,623 ± 466
10 ⁴	—	50 ± 21	817 ± 161	81 ± 13
	+	57 ± 2	424 ± 83	6,414 ± 437
10 ³	—	100 ± 51	104 ± 19	90 ± 43
	+	60 ± 7	161 ± 84	400 ± 161

DR-depleted cells and virus were prepared as in Fig. 1. Accessory cells autologous monocytes and B cells obtained after E-rosetting T-cell depletion of MNC were added in aliquots of 5 × 10⁵. Accessory populations did not proliferate in response to HTLV-I (data not shown). PHA was added at a final concentration of 1 µg ml⁻¹. Incorporation of ³H-dT was measured during the last 6 h of day 5 of culture. The ³H-dT incorporation of triplicate cultures is reported as mean ± s.d.

Furthermore, the transient proliferation of HTLV-I target cells could favour the integration of proviral DNA and the subsequent expression of the protein coded by the *x-lor* region of HTLV-I. The *trans*-acting transcriptional effects of this protein which lead to increased viral replication²⁴⁻²⁶, might also be related to the sustained proliferation of infected T lymphocytes²⁷. Activation induced by HTLV-I represents the main early event occurring after the attachment of viral particles to the lymphocyte membrane, so further understanding of the mechanisms implicated in this early proliferation should focus on the function of T-cell receptors and the viral envelope glycoproteins.

We thank Jan de Vries, Hans Yssel and Jacques Bertoglio for helpful comments, and Samuel Broder for 0.5α monoclonal antibody. This work was supported by grants from Fédération des Centres de Lutte contre Le Cancer (Vaincre le Cancer), Paris and from Association pour la Recherche sur le Cancer, Villejuif.

Received 12 February; accepted 4 March 1987.

- Weiss, A. *et al.* *A. Rev. Immun.* **4**, 593-619 (1986).
- Fitch, F. W. *Microbiol. Rev.* **50**, 50-69 (1986).
- Greene, W. C. & Leonard, W. J. *A. Rev. Immun.* **4**, 69-95 (1986).
- Wakasugi, H., Bertoglio, J., Tursz, T. & Fradelizi, D. *J. Immun.* **135**, 321-327 (1985).
- Katzen, D. *et al.* *J. Immun.* **135**, 1840-1845 (1985).
- Wong-Staal, F. & Gallo, R. C. *Nature* **317**, 395-403 (1985).
- Yamamoto, N. & Hinuma, Y. *J. gen. Virol.* **66**, 1641-1660 (1985).
- Kobayashi, N. & Hatanaka, M. *Cancer Rev.* **1**, 64-95 (1986).
- Franchini, G., Wong-Staal, F. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6207-6211 (1984).
- Duc Dodon, M. & Gazzolo, L. *Blood* **69**, 12-17 (1987).
- Popovic, M., Lange-Wantzin, G., Sarin, P. S., Mann, D. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5402-5406 (1983).
- Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
- Matsushita, S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 2672-2676 (1986).
- Gazdar, A. F. *et al.* *Blood* **55**, 409-417 (1980).
- Miyoshi, I. *et al.* *Nature* **294**, 770-771 (1981).
- Salahuddin, S. Z. *et al.* *Virology* **129**, 51-64 (1983).
- Wainberg, M. A., Spira, B., Boushira, M. & Margolese, R. G. *Immunology* **54**, 1-7 (1985).
- Uchiyama, T., Broder, S. & Waldmann, T. J. *Immun.* **126**, 1393-1397 (1981).
- Graneli-Piperno, A., Andrus, L. & Steinman, R. M. *J. exp. Med.* **163**, 922-937 (1986).
- Depper, J. M., Leonard, W. J., Robb, R. J., Waldmann, T. A. & Greene, W. C. *J. Immun.* **131**, 690-696 (1983).
- Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K. & Uchino, H. *Blood* **50**, 481-492 (1977).
- Kinoshita, K. *et al.* *Blood* **66**, 120-127 (1985).
- Matutes, E., Dalglish, A. G., Weiss, R. A., Joseph, A. P. & Catosky, D. *Int. J. Cancer* **38**, 41-45 (1986).
- Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. *Science* **225**, 381-385 (1984).
- Fujisawa, J. I., Seiki, M., Kiyokawa, T. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2277-2281 (1985).
- Cann, A. J., Rosenblatt, J. D., Wachsmann, W., Shah, N. P. & Chen, I. S. Y. *Nature* **318**, 571-574 (1985).
- Greene, W. C. *et al.* *Science* **232**, 877-880 (1986).
- Yssel, H., De Vries, J. E., Koken, M., Van Blitterswijk, W. & Spits, H. *J. Immun. Meth.* **72**, 219-227 (1984).
- Gillis, S., Ferm, M. M., Ou, W. & Smith, K. A. *J. Immun.* **120**, 2027-2032 (1978).

Evidence that the recently discovered θ 1-globin gene is functional in higher primates

Jeng-Pyng Shaw*, Jon Marks & Che-Kun James Shen

Department of Genetics, University of California, Davis, California 95616, USA

A new subfamily of the α -globin-like family has recently been identified in higher primates¹⁻³, rabbit⁴, galago⁵ and possibly the horse⁶. One member of this subfamily, θ 1, is downstream from the adult α 1-globin gene. In orang-utan, but not in rabbit⁴ or galago⁵, the θ 1-gene appears to be structurally intact, suggesting that it may be functional in this species². The orang-utan θ 1-gene possesses initiation and termination codons, and the predicted polypeptide differs from the orang-utan α 1-globin by 55 amino acids. The upstream promoter boxes CCAAT and ATA are present, although approximately 150 base pairs (bp) farther upstream than in the α 1-gene. This structural difference in the promoter between the orang-utan θ 1- and α 1-genes has led Proudfoot⁷ to speculate that the θ 1-gene may be inactive. We have now cloned the θ 1- and α 1-globin genes from the olive baboon, and have compared their sequences with those of orang-utan. The unique promoter structure of the orang-utan θ 1-gene is highly conserved in baboon, although the orang-utan and baboon diverged nearly 30 million years ago. The coding sequences of the two θ 1-genes differ by only 6.3% with 22 out of 27 nucleotide substitutions being codon third position silent changes. These data support the view that the θ 1-gene has been functional in the baboon, orang-utan, and by implication, in man. We also estimate that the duplication event generating the θ 1- and α -globin-like subfamilies may have occurred as much as 260 million years ago.

The olive baboon, *Papio anubis* (Primates: Cercopithecidae), is a close relative of the apes and humans, having diverged from that lineage nearly 30 million years ago (refs 8-11; Fig. 1a). The structure of its adult α -globin gene region is similar to that of the human and orang-utan. They all contain duplicated adult α -genes (α 2 and α 1), and a θ 1-gene located approximately 3 kilobases (kb) downstream from the α 1-globin gene (Fig. 1b). The α 1- θ 1 intergenic DNA of the orang-utan is 500 base pairs (bp) longer than either the human¹² or the baboon region (Fig. 1b), and the α 2- α 1 intergenic DNA of the baboon is 900 bp longer than the human and the orang-utan regions (Fig. 1b). The molecular basis of these length variations has not been characterized. From a baboon genomic library, we have isolated a recombinant phage, BαG1, containing the baboon α 1- and θ 1-genes³. The nucleotide sequences of these two genes and their flanking DNA, determined by the Maxam-Gilbert method¹³, are shown and compared to the orang-utan α 1- and θ 1-globin genes^{1,2} in Figs 2 and 3.

Although the adult α -globin genes, α 2 and α 1, are functional in human, orang-utan and baboon, no product of the neighbouring θ 1-gene has yet been identified. However, the orang-utan θ 1-gene has exon lengths identical to the functional α - and ζ -globin, has no premature termination codons, and has an apparent promoter, splicing sites and messenger RNA processing signals: all suggesting to us that the θ 1-gene is functional². Alternatively it may be an inactive pseudogene in the orang-utan as well as in the other higher primates. This possibility has been raised on the basis of the unique promoter structure of the orang-utan θ 1-gene, and the apparent lesions, including frame-shifts and absence of promoter sequences, in the orthologous θ 1-genes of the rabbit⁴, galago⁵ and possibly the horse⁶.

The two contrasting hypotheses, whether the θ 1-gene is or is

* Present address: Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA.

not functional in the higher primates, generate different predictions about the pattern of nucleotide substitutions between the orang-utan $\theta 1$ -gene and an orthologous gene from a closely related species, such as the $\theta 1$ -gene of the olive baboon. We have cloned this gene to examine the hypothesis that $\theta 1$ is a pseudogene in the catarrhine primate lineage. The sequence divergence between the two primate $\theta 1$ -genes, if they both have been functional, is expected to be lower for their exons and regulatory regions such as promoter sequences, than for their introns and other non-coding regions. Furthermore, if the $\theta 1$ -gene has been functional in the higher primates, we would expect that the majority of nucleotide substitutions between the orang-utan and baboon $\theta 1$ homologous genes would be silent mutations instead of replacement mutations. The comparison between the $\theta 1$ - and $\alpha 1$ -gene divergences of baboon and orang-utan should also reveal whether the $\theta 1$ -genes have been evolving at a rate roughly similar to the $\alpha 1$ -genes, which would imply similar selective constraints on both genes. If the rate of change of the $\theta 1$ gene is similar to that of other globin genes, then the approximate time at which the α - and $\theta 1$ -genes began to diverge can be calculated.

The baboon $\theta 1$ -globin gene contains no apparent defects, and is very similar in structure to the orang-utan $\theta 1$ -gene (Fig. 2). The lengths of the exons are identical between the two species, yielding a predicted protein of 141 amino acids without a premature termination codon. There is 6.3% divergence in the coding regions and approximately 8.6% divergence in the non-coding regions. The sequences near the intron-exon splice junctions and the promoter elements are particularly well conserved among the non-coding sequences. For example, from base -100 to the putative cap site of the $\theta 1$ -gene (+1, Fig. 2), which contains the CCAAT and TATA boxes, there are only three base differences between the orang-utan and baboon.

Most striking, however, is the distribution of base substitutions in the putative coding sequence of the $\theta 1$ -gene. Of 27 substitutions, only five are amino-acid substitution mutations, whereas 22 are silent changes in the third codon position. The number of replacement substitutions is that expected for functional α -globin-like genes¹⁴⁻¹⁷; the number of silent substitutions is much higher than expected for a non-functional gene. Given that the length of the θ -gene coding sequence is $3 \times 141 = 423$ bp, and that the Old World monkey versus hominoid divergence over non-coding DNA is about 8%, we would expect $423 \times 0.08 \times 0.33 = 11$ silent substitutions if the gene is non-functional. That this distribution is not artefactual can be demonstrated by organizing the intron sequences into triplets, and observing that the differences between orang-utan and baboon are distributed roughly at random among first, second and third positions. These data support the theory that evolution of the $\theta 1$ -gene is strongly constrained by natural selection in these higher primates, and that the gene performs some function, as yet unknown.

The evolutionary history of the primate $\theta 1$ -gene can be estimated from the percentage of replacement site substitutions

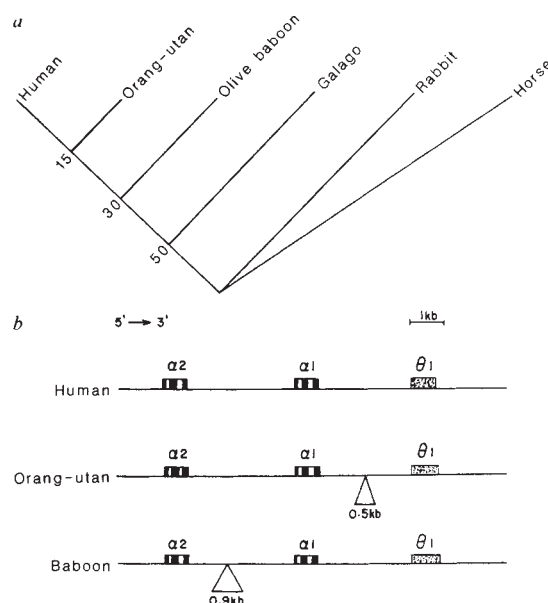


Fig. 1 a, Phylogenetic relationships of four primates (human, orang-utan, baboon, galago) and two non-primate mammals (rabbit, horse). The approximate times of divergence for the primate species⁸⁻¹¹ are given in millions of years. b, Alignment of the adult α -globin gene regions of human, orang-utan, and olive baboon. Triangles, length differences among the three species detected by Southern blot hybridization (J.-P. Shaw, data not shown). The difference between the orang-utan and human has been detected previously by Zimmer *et al.*¹². For the adult $\alpha 2$ - and $\alpha 1$ -genes, exons are in black and introns in white. Stippled boxes, $\theta 1$ -genes.

when compared with other α -globin-like genes. As shown in Table 1, pairwise comparisons of the primate $\alpha 1$ - and $\theta 1$ -genes, either within or between species, yield similar values for replacement site divergence, namely about 26%.

From Fig. 3, we can obtain an estimate of the genetic divergence between the orang-utan and baboon $\alpha 1$ -genes. Here we find 26 base differences in the coding regions, of which 12 are replacement and 14 are silent substitutions. The predicted amino-acid sequence is identical to the published α -globin sequence of *Papio cynocephalus* (*= P. anubis*)¹⁸. The $\theta 1$ -gene divergence compares favourably to the $\alpha 1$ -gene divergence data, with a similar number of differences, but distributed more conservatively. The non-coding regions of the $\alpha 1$ -gene diverge by 4.9%, as compared to 6.1% for the coding region. It should be noted, however, that the target size for the non-coding regions of the $\alpha 1$ -gene is considerably smaller than for the non-coding regions of the $\theta 1$ -gene. Further, the number of α -globin amino-acid replacements in the baboon lineage is known to be anomalously high¹⁹. We believe the evolution of the $\theta 1$ -gene compared to the $\alpha 1$ -gene does not yield the contrast expected

Table 1 Percentages of replacement-site and silent-site substitutions among the baboon $\alpha 1$, baboon $\theta 1$, orang-utan $\alpha 1$, orang-utan $\theta 1$, human α and human ζ globin genes

Human α	38 (60)						
Orang-utan $\alpha 1$	38 (60)	1.1 (6.6)					
Orang-utan $\theta 1$	42 (54)	26 (51)	26 (45)				
Baboon $\alpha 1$	41 (64)	5.5 (6.4)	6.4 (7.6)	28 (44)			
Baboon $\theta 1$	41 (68)	26 (55)	26 (52)	1.7 (22)	25 (50)		
Rabbit α	41 (60)	12 (29)	11 (33)	28 (55)	15 (30)	27 (61)	
Rabbit $\psi\alpha(\theta)$	44 (55)	30 (66)	39 (50)	14 (26)	33 (54)	13 (39)	32 (51)
Human ζ		Human α	Orang-utan $\alpha 1$	Orang-utan $\theta 1$	Baboon $\alpha 1$	Baboon $\theta 1$	Rabbit α

The numbers represent the percentages of replacement-site substitutions (with silent-site substitutions in parentheses) over different pairs of α - and $\theta 1$ -globin genes. Calculations were made according to the method of Perler *et al.*¹⁴. The comparison between human $\alpha 1$ ²⁹ and ζ -genes²⁶ have been made previously by Czelusniak *et al.*²⁴ and Proudfoot *et al.*²⁶. Also included is the comparison between rabbit α^4 , $\psi\alpha(\theta)^4$ genes and the primate globin genes (additional data from ref. 1).

Orang-utan $\theta 1$ ATCCAGCTA CTGGGAGGC TGAGGCAGGA GAATCGTTG -128
 Olive baboon $\theta 1$ AACCTGGAG GCGGAGTTG CCGTGAGCG AAATCGCGCC ACTGCACTCA CCGCACCCGG -68
 CCAATTTTGT TGTTTTGTAGT AGAGACTAAA AACTATATGG TGAACACCTA AGACGCGGGG -8
 GCCTTGGATC CAGGGCGATT CAGAGTTCCC CCGGTCGGAG T CAGTCGGAGA TGGAGGCCGC 53
 GCGGTCCCGG GATCCGGGAC CAGGCCCTGC ACCCCAGGGT GCGGAGGCTG CAGCGCGGGG 113
 CCCCCTGGTG GCGCGGGGAC CCGTGGCCGG TCCGCGCAGG GCGAGCGCGG GCGCAGGGCG 173
 CCGCGGGTTC CAGCGCGGGA ATG GCG CTG TCC GCG GAG GAC CGG GCG CTG GTG 226
 GGC GCC CTG TGG AAG AAA CTG GGA AGC AAT GTT GGC GTC TAT GCT ACT G 275
 Arg Ala Leu Trp Lys Lys Leu Gly Ser Asn Val Gly Val Tyr Ala Thr G
 AG GCC CTG GAG AG GTGCGGCGAG GCTGGGCGCC CCGCGCCTCC GGGGCCCTGC CTC 331
 lu Ala Leu Glu Ar
 CCAAGC CCGCGGAGC GCGCTCACCG CCGTTCCTCT CCGAG G ACC TTC CTG GCT 386
 g Thr Phe Leu Ala
 TTC CCC GCC ACG AAG ACC TAC TTC TCC CAC CTA GAC CTG AGC CCC GGC T 435
 Phe Pro Ala Thr Lys Thr Tyr Phe Ser His Leu Asp Leu Ser Pro Gly S
 CC GCC CAG GTT AGA GCA CAC GGC CAG AAG GTG GCG GAC GCG CTG AGC CT 484
 er Ala Gln Val Arg Ala His Gly Gln Lys Val Ala Asp Ala Leu Ser Le
 C GCC GTG GAG GCG CTA GAC GAC CTA CCC GCG GCG CTG TCC GCT CTG AGC 533
 u Ala Val Glu Arg Leu Asp Asp Leu Pro Arg Ala Leu Ser Ala Leu Ser
 CAT CTG CAC GGT TGC CAG CTG CGA GTG GAC CCA GCT AAC TTC CCG GTGAG 583
 His Leu His Ala Cys Gln Leu Arg Val Asp Pro Ala Asn Phe Pro
 GCGCT GCGGTGCTG GCCCTGTTC TCGGGAGGGC TTGCGCGGGG CCGGTGCGGG GGGC 642
 G CAC GCGGGGTGCA GGCAAGTGAG CTTTGGCGCG TCGCCGCGAG CTC CTG GGC CAC 699
 ATGTGG GCGGGGTGCA GGCAAGTGAG CTTTGGCGCG TCGCCGCGAG CTC CTG GGC CAC
 Leu Leu Gly His
 TGC CTG CTG GTG ACC CTC GCC CCG CAC TAC CCC GGA GAC TTC AGC CCG G 748
 Cys Leu Leu Val Thr Leu Ala Arg His Tyr Pro Gly Asp Phe Ser Pro A
 CG CTG CAG GCG TCG CTG GAC AAG TTC CTG AGC CAC GTG ATC TCT GCG CT 797
 la Leu Gln Ala Ser Leu Asp Lys Phe Leu Ser His Val Ile Ser Ala Le
 G GCT TCC GAG TAC GCG TGA ACTGTGGGTG GGTGGCCGTG TGACCCCGAG GCGATC 852
 u Ala Ser Glu Tyr Arg TER
 TTCC CCGTGTATGA GTAAAGCCTT CCGCAAGGTGC AGCCTTCTTG CCGTGTCTC TCAAA 911
 GGCAG GACGCGAGAA CAAGGCGCCG CCGCGCCCCA AAGGA 951

Fig. 2 Sequence comparison of the baboon and orang-utan $\theta 1$ -globin genes. Nucleotide sequence of the baboon gene is given in full. Amino-acid translation is given below the presumptive coding sequences. Above the baboon sequences, nucleotides and amino acids of the orang-utan $\theta 1$ -gene which differ from the baboon $\theta 1$ -gene and its flanking DNA: asterisk, deleted base. Promoter boxes (CCAAT and TATA), initiation codon, termination codon, polyadenylation signal, and polyadenylation site are all underlined. Numbers, base positions relative to the putative cap site (+1) (arrow). Orang-utan sequence from ref. 2.

between pseudogenes and functional genes.

Assuming that the globin genes tend to accumulate amino-acid substitution mutations at an average rate of 1% change per 10 million years¹⁴⁻¹⁷, we estimate from Table 1 that the $\theta 1$ -gene began to diverge from the adult α -globin genes approximately 260 million years ago. The divergence between various $\alpha 1$ -globin genes and the human (embryonic) ζ -globin gene is nearly the same as that between the human ζ -globin gene and the two anthropoid θ -globin genes, confirming the ancient history of the $\theta 1$ -globin gene.

We have assumed a constant average rate of molecular evolution²⁰⁻²³ of the $\theta 1$ - and α -globin genes. However, this rate

Orang-utan $\alpha 1$ CCAATGAGCA GCGCCCTGCC GGGCGTGCCC CCGCGCCCGG -32
 Olive baboon $\alpha 1$ AGCATAAACC CTGGCGCGCT CCGCGCCCGG CACTCTTCTG GTCCCCACAG ACTCAGAAAG 29
 AACCACCC ATG GTG CTG TCT CCT GAC GAC AAG AAA CAC GTC AAG GCC GCC 79
 INI Val Leu Ser Pro Asp Asp Lys Lys His Val Lys Ala Ala
 TGG GGT AAG GTC GGC GAG CAC GGT GGC GAG TAT GGT GCG GAG GCC CTG G 128
 Trp Gly Lys Val Gly Glu His Ala Gly Glu Tyr Gly Ala Glu Ala Leu G
 AG AG GTGAGGCTCC CTCCTCTGCT CCGGCGCGGG CTCTCGCCC ***** ***G 186
 lu Ar
 GCCACC CGCAACCGTC CTGGCCCGGG ACCCAAACCC CACCCTCAC TCTGTTTCTC CCC 245
 GCAG G ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC TTC CCC CAC TT 294
 g Met Phe Leu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His Ph
 C GAC CTG AGC CAC GGC TCT GAC CAG GTT AAC AAA CAC GGC AAG AAG GTG 343
 e Asp Leu Ser His Gly Ser Asp Gln Val Asn Lys His Gly Lys Lys Val
 GCC GAC GCG CTG ACC CTC GCC GTG GGG CAC GTG GAC GAC ATG CCC CAG G 392
 Ala Asp Ala Leu Thr Leu Ala Val Gly His Val Asp Asp Met Pro Gln A
 OG CTG TCC AAG CTG AGC GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC CC 441
 la Leu Ser Lys Leu Ser Asp Leu His Ala His Lys Leu Arg Val Asp Pr
 G GTC AAC TTC AAG GTGAGCGGCG GCGCGGAGC GATCTGGTTC GAGGGCGAG ATG 497
 o Val Asn Phe Lys
 GCGCCTT CCTCGCAGGG CACAGCATCC GCGAGTTGC GGGAGGTGCA GCGCAGGCGG CG 556
 GCTGCGGG CCGTCCGGCC CACTGACCTT CTCTCTGCA CAG CTC CTG AGC CAC TGC 612
 Leu Leu Ser His Cys
 CTG CTG GTG ACT CTG GCC GGT CAC CTC CCC GCG GAG TTC ACC CCT GCG G 661
 Leu Leu Val Thr Leu Ala Ala His Leu Pro Ala Glu Phe Thr Pro Ala V
 TG CAC GCC TCC CTG GAC AAG TTC CTG GCT TCT GTG AGC ACC GTG CTG AC 710
 al His Ala Ser Leu Asp Lys Phe Leu Ala Ser Val Ser Thr Val Leu Th
 C TCC AAA TAC CGT TAA GGTGGAGCCT CCGTGGCCAT GCTTCTTGCC CCGTGGCCCT 766
 r Ser Lys Tyr Arg TER
 CCGGCCAGGC CCTCCTCCCC TTCTGTCACC CGTACCCCC* TGGTCTTTGA ATAAATG 823

Fig. 3 Sequence comparison of the baboon and orang-utan $\alpha 1$ -globin genes. The complete sequence of the baboon is given, with translation below the coding sequences. For the orang-utan sequence, only nucleotides and amino acids which differ from the baboon are given. Asterisks, deleted bases. (There is a 13-bp deletion in intron 1 of the baboon $\alpha 1$ -gene which we interpret as a single mutational event.)

may not be constant^{24,25}. An alternative is that after the duplication separating the $\theta 1$ -gene from the α , the $\theta 1$ evolved faster than the α -gene. If this is the case, then the duplication event generating the α - and $\theta 1$ -globin genes may have occurred more recently than 260 Myr ago. This possibility is supported by parsimony analyses²⁵ which may place the α - θ separation at 110 Myr ago (B. Koop and M. Goodman, personal communication).

The minor features which appear to render the rabbit orthologue of $\theta 1$ an inactive pseudogene are probably of recent origin⁴. From the percentages of replacement site changes and the neutral substitutions (Table 1) between the orthologous rabbit $\psi\alpha(\theta)$ and orang-utan $\theta 1$ -gene, or the baboon $\theta 1$ -gene, we can estimate, using the method of Proudfoot and Maniatis²⁶, that the rabbit $\psi\alpha(\theta)$ became inactivated approximately 15 million years ago. An interesting question is why the structure of the orthologous $\theta 1$ -genes of the galago (a strepsirrhine, prosimian, or lower primate) and rabbit seem to preclude the possibility of their expression. One possibility is that the novel form of placental evolution by the higher (haplorhine) primates^{27,28} places demands for globin synthesis on the genetic

system to which the retention of a functional $\theta 1$ -globin gene is an adaptive response. Another possibility is that the gene also exists in multiple copies in the galago and rabbit, as in the baboon, orang-utan and human², and that although the orthologue of the $\theta 1$ -gene may be silenced, another as yet undetected copy may be functional.

We believe that the data presented here indicate strong selective constraints on the evolution of the catarrhine $\theta 1$ -gene. This is most compatible with the hypothesis that it is indeed functional in some tissues at a particular developmental stage in baboons, orang-utans, and by implication, in humans. This inference is supported independently by the existence of a processed θ -globin pseudogene fragment, containing part of exon 3, an intact 3'-untranslated region and poly(A) tail, found on chromosome 22 in the human genome³⁰.

We thank Ikuhisa Sawada, John Gillespie and Celia Shen for quantitative help, Carl Schmid for access to unpublished data. We thank the California Regional Primate Center for baboon tissue, and Kimberly Strauch for typing the manuscript. This research was supported by a grant from NIH, a grant from the American Cancer Society, and a Research Career Development Award to C.-K. James Shen. C.-K.J.S. is on sabbatical at the Institute of Molecular Biology in Taiwan, ROC, under the support of the National Council of Science, ROC, and a NIH-Fogarty International fellowship. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00267.

Estimation of wedge components in curved DNA

Levy E. Ulanovsky & Edward N. Trifonov

Department of Polymer Research, Weizmann Institute of Science, Rehovot 76100, Israel

There are growing indications that the inherent curvature of DNA is important in protein-DNA recognition (for reviews see refs 1-3). The 10.5-base-pair (bp) periodicity of some dinucleotides first found in eukaryotic DNA sequences^{4,5} was interpreted as the expression of curvature of periodic segments of double-stranded DNA, the curvature resulting from co-orientation of periodically spaced 'wedges' between stacked base pairs. The wedge can be decomposed into roll and tilt components, opening towards a groove and a backbone respectively (Fig. 1a), both contributing to DNA curvature (Fig. 1b and c). The largest wedge was estimated to belong to the AA·TT dinucleotides. Recent work^{6,7} provided new experimental data on synthetic curved DNA. The authors tried to apply the wedge model to their results and met problems in doing so. We have found that taking into account both roll and tilt components of the AA·TT wedge, in the correct ratio, leads to remarkable consistency between the wedge model and the data.

The inherent curvature of free double-stranded DNA can be conveniently observed as an anomaly in its migration in PAGE⁸ (polyacrylamide gel electrophoresis), where the curved fragments presumably encounter higher friction. All electrophoretically anomalous DNA fragments found so far exhibit a 10- to 11-bp periodicity in the distribution of AA and TT dinucleotides in their sequence¹. The absolute curvature of DNA can be measured by a non-constrained circularization technique⁹.

An important observation was made recently⁷ that a tandemly repeated decamer (NAAAATTTTN)_n showed highly anomalous migration in PAGE, but (NTTTTAAAN)_n did not (N represents C or G). The author applied symmetry considerations, arguing that the dinucleotide wedge model should predict equal curvature for the two repeats and thus could not be valid.

Received 1 December 1986; accepted 11 February 1987.

1. Marks, J., Shaw, J. P. & Shen, C.-K. *J. Proc. natn. Acad. Sci. U.S.A.* **83**, 1413-1417 (1986).
2. Marks, J., Shaw, J. P. & Shen, C.-K. *J. Nature* **321**, 785-788 (1986).
3. Marks, J. *et al. Cold Spring Harb. Symp. quant. Biol.* (in the press).
4. Cheng, J.-F., Raid, L. & Hardison, R. C. *J. biol. Chem.* **261**, 839-848 (1986).
5. Sawada, I. & Schmid, C. W. *J. molec. Biol.* **192**, 693-709 (1986).
6. Clegg, J. B., Goodbourn, S. E. Y. & Braend, M. *Nucleic Acids Res.* **12**, 7847-7858 (1984).
7. Proudfoot, N. *Nature* **321**, 730-731 (1986).
8. Gingerich, P. & Schoeninger, M. *J. hum. Evol.* **6**, 483-505 (1977).
9. Fleagle, J. in *Major Topics in Primate and Human Evolution* (eds Wood, B. A., Martin, L. & Andrews, P.) 130-149 (Cambridge University Press, 1986).
10. Sibley, C. & Ahlquist, J. *J. molec. Evol.* **20**, 2-15 (1984).
11. Andrews, P. *Nature* **314**, 398-499 (1985).
12. Zimmer, E. A., Martin, S., Beverley, S., Kan, Y. W. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2158-2162 (1980).
13. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
14. Perler, F. *et al. Cell* **20**, 555-566 (1980).
15. Proudfoot, N. & Maniatis, T. *Cell* **21**, 537-544 (1980).
16. Li, W.-H., Luo, C.-C. & Wu, C.-I. in *Molecular Evolutionary Genetics* (ed. MacIntyre, R. J.) 1-84 (Plenum, New York, 1985).
17. Efstradiatis, A. *et al. Cell* **21**, 653-668 (1980).
18. Mahoney, W. & Nute, P. *Biochemistry* **19**, 1529-1534 (1980).
19. Hewett-Emmett, D. *et al. in Molecular Anthropology* (eds Goodman, M. & Tashian, R.) 257-275 (Plenum, New York, 1976).
20. Dickerson, R. E. *J. molec. Evol.* **1**, 26 (1971).
21. Langley, C. H. & Fitch, W. M. *J. molec. Evol.* **3**, 161-177 (1974).
22. Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* **46**, 573-639 (1977).
23. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, London 1983).
24. Czelusniak, J. *et al. Nature* **298**, 297-300 (1982).
25. Goodman, M. *Prog. Biophys. molec. Biol.* **38**, 105-164 (1981).
26. Proudfoot, N., Gil, A. & Maniatis, T. *Cell* **31**, 553-563 (1982).
27. Luckett, W. P. in *Phylogeny of the Primates: A Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 157-182 (Plenum, New York, 1975).
28. Martin, R. D. in *Phylogeny of the Primates: A Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 265-297 (Plenum, New York, 1975).
29. Liehaber, S. A., Goossens, M. J. & Kan, Y. W. *Nature* **290**, 26-29 (1981).
30. Shaw, J.-P. *et al. in Hemoglobin Switching V* (ed. Stamatoyannopoulos, G., in the press).

In fact, however, there is no symmetry between 5'-TTTTAAAA-3' and 5'-AAAATTTT-3' because of the different symmetry properties of roll and tilt components of the AA·TT wedge. In replacing a dinucleotide by its complementary dinucleotide, the tilt reverses its direction, but the roll does not. Depending on the relationship between roll and tilt in the AA·TT wedge, either one of TTTTAAAA or AAAATTTT can have zero overall curvature, the other one being significantly curved. The reason for this effect can be visualized by considering a series of simple 10-bp repeats, with the repeat unit containing one AA and one TT dinucleotide (Fig. 2). The long and short vectors in Fig. 2 represent rolls and tilts correspondingly. The normally migrating (NTTTTAAAN)_n can also be written as (AAAANNTTTT)_n and, thus, corresponds to Fig. 2e. The zero vectorial sum in Fig. 2e illustrates that the six AA·TT wedges in the AAAANNTTTT·AAAANNTTTT double helix can cancel each other:

$$r \cos(a/2) - t \sin(a/2) = 0 \quad (1)$$

where a (144° for the assumptions of Fig. 2e) is the helical angular separation between the middles of the adjacent runs of TTTT and AAAA (in terms of base pairs $a = 10 - 6 = 4$ bp), and t and r are tilt and roll components of the small AA·TT wedge.

The other repeat, NAAAATTTTN·NAAAATTTTN, corresponds to Fig. 2c, and exhibits⁷ highly anomalous migration. The curvature of this DNA is proportional to

$$r \cos(a/2) + t \sin(a/2) \neq 0 \quad (2)$$

where the left hand side differs from that of (1) only in the sign of the second term. The change results from reversing the sign of the angular separation between the AAAA and TTTT, which leads to a coherent contribution of the two terms in (2) to the overall curvature in Fig. 2c, in contrast to Fig. 2e.

From (1) we derive the estimate $r/t = 3.1$, of the roll-to-tilt ratio in the AA·TT minihelix. A similar but more refined computation, which takes into account the sequence dependence of helical twists¹⁰ and the local curvature within each AAAA and TTTT run, yields an r/t ratio of ~ 3.5 . The total AA·TT wedge angle (roll and tilt combined) was recently measured⁹ by a non-constrained circularization technique and found to be 8.7° . The conditions $r^2 + t^2 = 8.7^2$ and $r/t = 3.5$, are sufficient to calcu-

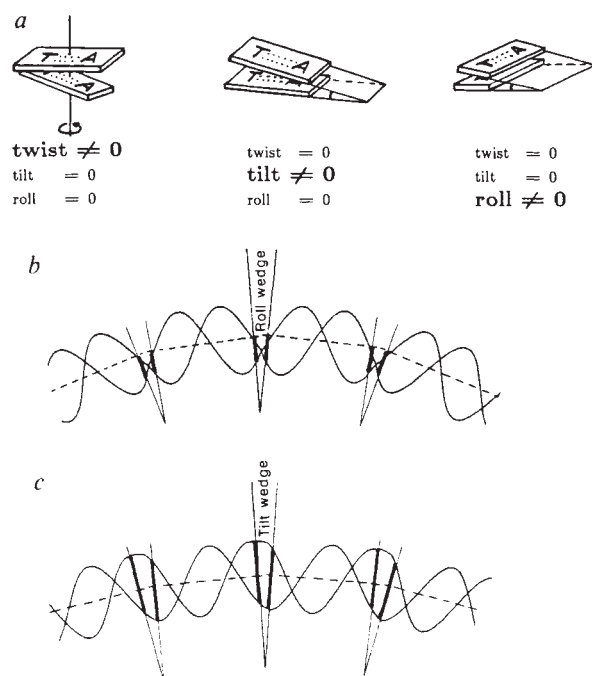


Fig. 1 Tilt and roll angles. *a*, Twist, tilt and roll angles formed by two adjacent base pairs. *b*, Curvature by roll components of the wedges, opening towards the major groove. *c*, Curvature by tilt components of the wedges, opening towards the backbone. Note that *b* and *c* show mutually perpendicular projections of the same DNA fragment containing three wedges separated by one helical turn (here 10 bp), thus causing unidirectional curvature of DNA. Tilts in *b* and rolls in *c* are not seen, being perpendicular to the plane of the paper.

late the previously unknown values of roll and tilt in the AA·TT wedge: $r = 8.4^\circ$ and $t = 2.4^\circ$. These two quantities are essential for computing the shape of any DNA fragment curved by AA·TT wedges. Without the proper values of the two components, given the total wedge, the calculated shape can only be correct for DNA with all AA dinucleotides in one strand and all TT dinucleotides in the other (for example, the plots in ref. 7). Figure 3 shows computer-generated diagrams of DNA molecules corresponding to expressions (1) and (2). The importance of the correct combination of roll and tilt components becomes clear if it is realized that, should either roll or tilt component alone be present in the AA·TT wedge, Fig. 3*a* and Fig. 3*b* would show exactly the same curvature.

The normal migration of 5'-NTTTTAAAN-3' determines also the relative directions of the two wedge components. Reversing the direction of either roll or tilt component of the AA·TT wedge interchanges Fig. 3*a* and *b*, whereas reversing both roll and tilt leaves the curvatures unchanged. The electrophoretic behaviour of AAAA·TTTT and TTTT·AAAA corresponds to an AA·TT wedge in which either the roll opens towards the major groove and the tilt does so towards the AA dinucleotide, or alternatively, the roll opens towards the minor groove, and the tilt does so towards the TT dinucleotide.

The two 20-bp repeats⁶ GCAAAAACGGCAAAAACG and GCAAAAACGGGTTTTTTGG differ by the replacement of half the AA dinucleotides by TT dinucleotides reversing tilts but not rolls. The first repeat is curved by rolls and tilts combined, but the second one is curved by rolls alone, the tilt components cancelling each other. According to the wedge model with $r/t = 3.5$ as above, the ratio of the curvatures of the two fragments is

$$\sqrt{(t^2 + r^2)/r^2} = 1.05 \quad (3)$$

so the second fragment should be only slightly less anomalous than the first, as was in fact observed⁶.

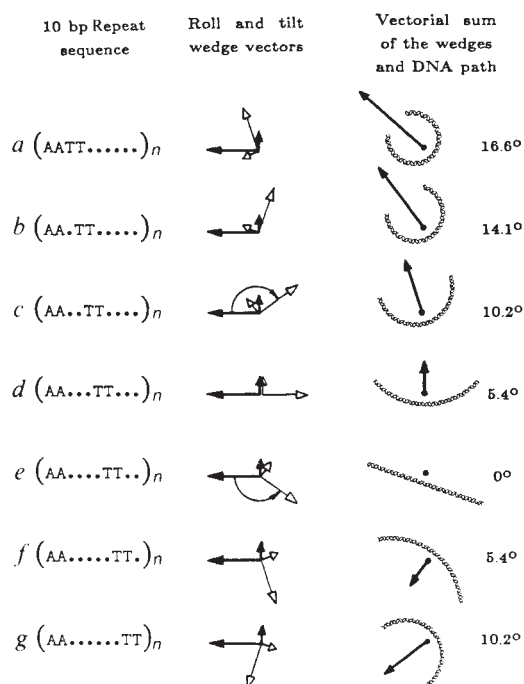


Fig. 2 Curvature caused by interplay of AA and TT wedges in a 10-bp repeat. Separating TT from AA by one more base results in a 36° rotation of TT versus AA wedge components denoted by unfilled (TT) and filled (AA) arrowheads in the central column, as viewed along the axis of the DNA. Each wedge component is shown as a vector pointing in the direction of its opening, the length of the vector being proportional to the opening angle. The long vectors are rolls, the short vectors are tilts. The numbers on the right are the magnitudes of the vectorial sum of AA and TT wedges of the central column, this sum being also the magnitude of the DNA axis deflection angle per 10 bp. In line *d*, the parallel and antiparallel orientations of tilts and rolls respectively, result from the 5-bp separation between AA and TT. The DNA pitch of 10.0 bp and equal helical twist for all base steps are assumed for simplicity. In line *e*, schematically corresponding to equation (1), the curvature (vectorial sum of wedges) is zero. The total AA·TT wedge angle $\sqrt{r^2 + t^2}$ (roll and tilt combined) is set to 8.7° as measured in ref. 7. The DNA path is computed with the algorithm described in Fig. 3.

The degree of anomaly in the PAGE migration of a fragment (measured as relative deviation of apparent from actual size) cannot be assumed *a priori* to be proportional to its curvature. Experiments indicate (S. Diekmann, personal communication; L.U. and P. Bucher, unpublished) that the gel matrix does not distinguish between straight and nearly straight molecules, the curvature threshold depending on the gel density. In fact, the symmetry of the function relating the anomaly to curvature implies that it has to be non-linear, perhaps quadratic at small curvatures. Slightly curved fragments migrate disproportionately close to normal as do, for example⁶, 10-bp repeats containing two non-adjacent AA dinucleotides.

This disproportionality⁶ was probably increased by the fact that a lower content of AA dinucleotides, which possess higher than average helical twist angles¹⁰, raises the pitch value farther out of register with the sequence repeat of 10 bp. The pitch influence on the migration anomaly was noted before⁶ for strong curvatures. Figure 4 shows that the pitch effect alone might provide a reasonable explanation for the different migration of otherwise equally curved repeats, which was considered a major problem for the wedge model⁶.

There are still many unresolved questions about both the wedge model and its alternatives. These include whether the junction bend model^{11,12} can be modified to explain the existing data (such as given in refs 7 and 13) as simply as the wedge model does. Concerning the wedge model, how significant are

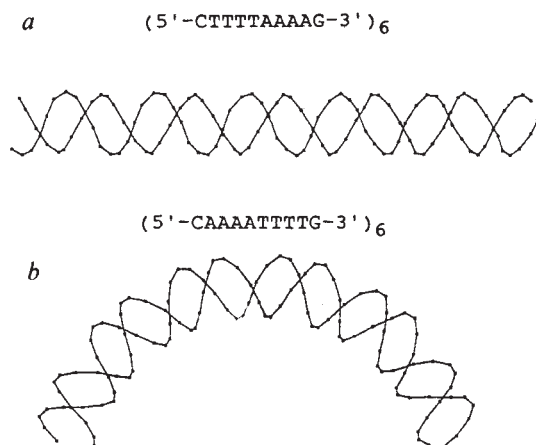


Fig. 3 Computer-generated diagrams of DNA paths of seemingly symmetrical but electrophoretically different repeats⁷: *a*, the $(\text{CTTTTAAAG})_6$ repeated decamer is electrophoretically normal; *b*, the $(\text{CAAATTTTG})_6$ repeated decamer is anomalously slow in PAGE. In computation of both *a* and *b* the roll angle of the AA·TT wedge is set to 8.4° opening towards the major groove and the tilt angle to 2.4° , opening towards the AA dinucleotide. Reversing both roll and tilt does not change the absolute curvature. The algorithm for computation of the phosphate coordinates is similar to the one in ref. 11 and involves helical translations by base steps with equal rise per residue and three sequence-dependent parameters for each of the 10 possible combinations of stacked base pairs: helical twist angle¹⁰; roll and tilt wedge angles. The latter are set to zero for dinucleotides other than AA and TT.

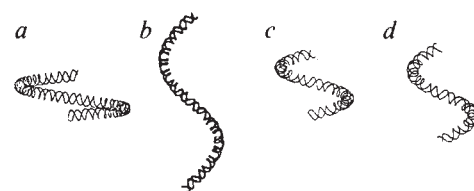
the contributions of dinucleotides other than AA·TT? To what extent does a given wedge angle depend on the flanking sequences: does an isolated AA·TT have the same wedge as an AA·TT within an oligo(A) run?

Of direct relevance to this last question is the observation that the migration anomalies of $(\text{GAAAATTTTC})_n$ and $(\text{GGAAATTTTC})_n$ fragments of up to about 100 bp are equally strong, although they differ slightly at longer lengths¹³. The unique roll-to-tilt ratio, which at short lengths makes the paths of the two molecules indistinguishable (both looking like the one in Fig. 3*b*), is $r/t = 3.5 \pm 0.5$, in agreement with the above estimate. Regarding the longer fragments, above 100 bp the two computed DNA paths turn out slightly different (Fig. 4*c* and *d*) because of the pitch effect discussed above, again remarkably consistent with experiment¹³. The significance of this consistency goes beyond another check of the wedge model and of the r/t ratio. It implies that a single AA·TT wedge within an AAAA run is equivalent to one within AAA—an important result *per se*. This is the only experiment known to us which compares the curvature of AAAA with that of a shorter oligo(A) run, but avoids the effect of the nonlinearity of the anomaly as a function of curvature, thanks to the comparable anomalies of the two fragments.

Regarding a comparison of the wedge and junction bend models, it was spectacularly shown by Prunell *et al.*¹⁴, that, if the DNA curvature is induced by stacking of wedges, it can also be viewed as caused by effective 'junction breaks' at the edges of the wedge stack, although no breaks actually occur at the junctions. The two models may turn out to be equivalent descriptions of the same phenomenon.

In contrast to the free DNA in solution discussed above, DNA complexed with proteins is likely to have different roll-to-tilt ratio. For example, the neutralization of the charges of the DNA phosphates facing histones should reduce their electrostatic repulsion. This can open an additional tilt wedge outwards from the core, so that in the nucleosome the tilt might perhaps be more important.

We thank J. Beckmann, P. Bucher, P. Davis and H. Eisenberg



Repeat sequence	CAAAAAAAAA	GGCCAAACCG	GAAAATTTTC	GGAAATTTCC
Sum of wedges	17°	17°	27°	27°
Pitch (bp)	10.17	10.45	10.32	10.44
Retardation in gel at 170 bp	1.9	1.3	2.4	2.2

Fig. 4 DNA pitch effect on migration. *a*, *b*, Two recently studied⁶ repeats drawn by computer (the algorithm is described in Fig. 3) showing a striking difference in shape (superhelical path), in spite of the equal number of uncanceled AA·TT wedges (vectorial sum). The effect is due to the slight difference in the pitch caused by the sequence dependence of the helical twist angles. The more the DNA pitch deviates from the sequence repeat length (10.0 bp here), the greater is the elongation of the DNA superhelix. Note that this non-constrained DNA is supercoiled by the inherent curvature only. For our program, the twist angles were taken from ref. 10 (the only values available), where numerous experimental data were weight-averaged, providing a set of all 10 twist angles. *c*, *d*, Repeats of ref. 13. The sum of wedges is the vectorial sum of the AA·TT wedges per 10-bp repeat. The retardation in gels is shown as the apparent size/actual size ratio in PAGE for 170-bp long fragments^{6,13}. The sum of wedges is the same for *a* and *b* and separately for *c* and *d*, but the migration anomaly is different, apparently due to the different elongation of the superhelix.

for valuable discussions. This work was supported in part by the Minerva Foundation.

Note added in proof: The 169-bp *tyrT* DNA (Drew and Travers, *J. molec. Biol.* **186**, 773–790) has been computer-graphed and found appreciably curved exactly in the plane of its circularization. The direction is correct only if the AA·TT roll opens to the major groove and the tilt does to AA, thus solving the remaining ambiguity in the wedge sign.

Received 22 October 1986; accepted 3 February 1987.

1. Trifonov, E. N. *CRC crit. Rev. Biochem.* **19**, 89–106 (1985).
2. Eisenberg, H. *Trends biochem. Sci.* **11**, 350–351 (1986).
3. Lilley, D. L. *Nature* **320**, 487–488 (1986).
4. Trifonov, E. N. & Sussman, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3816–3820 (1980).
5. Trifonov, E. N. *Nucleic Acids Res.* **8**, 4041–4053 (1980).
6. Koo, H. S., Wu, H. M. & Crothers, D. M. *Nature* **320**, 501–506 (1986).
7. Hagerman, P. J. *Nature* **321**, 449–450 (1986).
8. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7664–7668 (1982).
9. Ulanovsky, L. E., Bodner, M., Trifonov, E. N. & Choder, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 862–866 (1986).
10. Kabsch, W., Sander, C. & Trifonov, E. N. *Nucleic Acids Res.* **10**, 1097–1104 (1982).
11. Levene, S. D. & Crothers, D. M. *J. Biomolec. Struct. Dynam.* **1**, 429–435 (1983).
12. Wu, H. M. & Crothers, D. M. *Nature* **308**, 509–513 (1984).
13. Hagerman, P. J. *Biochemistry* **24**, 7033–7037 (1985).
14. Prunell, A., Goulet, L., Jacob, Y. & Goutorbe, F. *Eur. J. Biochem.* **138**, 253–257 (1984).

Erratum

Structure of D-ribulose-1,5-bisphosphate carboxylase/oxygenase from *Alcaligenes eutrophus* H16

A. Holzenburg, F. Mayer, G. Harauz, M. van Heel, R. Tokouka, T. Ishida, K. Harata, G. P. Pal & W. Saenger
Nature **325**, 730–732 (1987).

THIS letter was published with an incorrect title. The above title is correct. The co-authors K. Harata and G. P. Pal are at the Institut für Kristallographie der Freien Universität Berlin, Takustrasse 6, D-1000 Berlin 33, FRG.

Labelling nucleic acids for hybridization

M.W. Cunningham and C.R. Mundy

The enormous utility of labelled nucleic acid probes has encouraged the development, over the past decade, of many efficient and reliable labelling methods.

THE selection of an appropriate combination of label, labelling method and detection system for a particular experiment depends mainly on the required levels of sensitivity and resolution. For example, detection of single-copy human genes on Southern blots requires high sensitivity, but in this application resolution is not usually a priority. Conversely, localization of genes on chromosomal regions by *in situ* hybridization requires exquisitely fine resolution.

When determining the appropriate labelling method, the constraints of the precise hybridization application limit the number of choices. In cases where an intact probe of defined length is required, as in *S₁* mapping¹, an end-labelling method is preferable, and usually provides adequate sensitivity. For most filter and *in situ* hybridization applications, uniform labelling methods, which offer higher label density, are generally used. The most common uniform labelling methods are nick translation², random primer labelling³ and transcription from cloned promoters using RNA polymerases from bacteriophages such as SP6, T7 or T3⁴. The main characteristics of these labelling reactions are summarized in Table 1.

Uniform labelling

Although the properties of the hybridization probes produced in uniform labelling reactions are broadly similar, significant differences do exist between the products of the uniform labelling reactions. For example, most nick translation protocols produce significantly more probe than do other uniform labelling methods. This makes nick translation particularly suitable when carrying out multiple hybridization reactions, or when a high probe concentration is desired. The high probe yield also makes nick translation particularly appropriate when non-radioactive biotinylated probes are used because, in contrast to radioactive probes, these do not give elevated nonspecific binding when used at high concentrations.

Random primer labelling is relatively insensitive to the purity of the substrate DNA, and thus both miniprep and gel-purified fragments can be labelled efficiently with this technique³. It is therefore particularly useful for labelling restriction fragments that have been gel-purified to

remove potentially cross-hybridizing vector sequences. The reaction leads to net synthesis, which results in the incorporation of labelled nucleotide at concentrations above the level required for complete replacement of the equivalent nucleotide in the template. The amount of substrate DNA is usually low, and the labelled nucleotide can be of high specific activity. Consequently, a high specific activity probe is produced, which can facilitate the rapid detection of rare sequences.

Probes of high specific activity are also produced by RNA polymerase-based transcription. In this technique, the

colour selection of recombinants and single-strand origins of replication for sequencing, have been developed in recent years.

In general, the choice of uniform labelling method will not significantly affect the final level of resolution. The effect of the uniform labelling method on attainable sensitivity is also small compared to that of the label itself. It is very important to tailor the choice of label, whether radioactive or non-radioactive, to the sensitivity and resolution required.

Properties of labels

When radionuclides are used as labels for

Table 1 Major properties of uniform nucleic acid labelling methods

Template	Nick translation dsDNA	Random primer ssDNA (and denatured dsDNA)	Phage polymerase dsDNA
Optimal form	Linear or circular	Linear	Linear
Amount of template	0.5–1 µg	25ng	1–2 µg
Reaction time	~2 hours	~3 hours	~1 hour
Efficiency of incorporation	~60%	~75%	~75%
Nature of probe	DNA	DNA	RNA
Competing strand	Yes	Yes	No
Amount of probe	0.5–1 µg	~50ng	~250ng
Potential specific activity of probe (d.p.m per µg)	2×10^9	5×10^9	5×10^9
Probe length	~500bp (medium variance)	~200bp (high variance)	Defined
Insert specific	No	No	Yes
Requirement for subcloning	No	No	Yes
Labelling in agarose	(Yes)*	Yes	No

Values quoted are for commonly used protocols.

*Although effective labelling in agarose has been observed, its efficiency is variable.

single-stranded RNA is not denatured prior to hybridization and is therefore not effectively diluted by denatured template. In some cases, RNA probes show higher sensitivity than equivalent nick translated probes in both Northern blots⁵ and *in situ* hybridization⁶. This is most probably due to the comparatively high stability of RNA:RNA hybrids, and to the fact that probe reannealing does not occur. The use of RNA probes also allows the removal of nonspecifically bound probe by treatment with ribonuclease A. One possible disadvantage of RNA polymerase methods is the requirement to subclone the sequence which is to be labelled into a vector containing appropriate phage polymerase promoters. However, a variety of general purpose phage and plasmid transcription vectors, with additional features such as

autoradiography, the sensitivity and resolution achieved is directly related to the physical properties of the radionuclides. The level of sensitivity which can be obtained depends both on the attainable specific activity of the label and on its detection efficiency.

The attainable specific activity is very high for ³²P, high for ³⁵S and ¹²⁵I and relatively low for ³H. Detection efficiency depends upon the type and energy of radioactive emission from the label, and varies according to autoradiographic conditions. The level of resolution achievable is also determined by the emission type and energy of the radionuclide, because this determines the extent of silver grain production in the overlying photographic emulsion.

³²P emits highly energetic β-particles,

with a long mean path length, producing silver grains over a relatively large area of both photographic emulsion layers of double-sided film. Therefore, the resolution of ^{32}P labels is low. Detection efficiency and hence sensitivity is particularly high if an intensifying screen is used to generate light from β -particles that pass through the film. The lower energy β -particles emitted by ^{35}S can be best detected using a dense photographic emulsion, but cannot penetrate through the plastic backing of standard X-ray film to the opposite emulsion, and internal absorption within the sample itself can also occur. The sensitivity of ^{35}S is lower than that of ^{32}P , but resolution is correspondingly higher.

^{125}I emits both γ -radiation and electrons. The former readily penetrates film without efficient silver grain production and is best detected using an intensifying screen. The low-energy Auger electrons are stopped entirely within the emulsion. Thus high sensitivity may be obtained with a screen and high resolution without. The very weak β -emission of ^3H yields the highest resolution, but its low maximum specific activity and low detection efficiency necessitate long exposure times.

The most common non-radioactive label, biotin, is usually detected by a colour-generating streptavidin-enzyme complex. Under ideal conditions it can combine good sensitivity and resolution with rapid development time. However, for high sensitivity blot hybridizations, it is generally less robust than ^{32}P , which remains the label of choice. Other labels such as ^{35}S or direct enzyme conjugates⁷ can be used for lower sensitivity blotting applications, but their use is still relatively rare.

In-situ hybridization

Increasingly, ^{35}S and even ^{32}P are being used for *in situ* hybridization^{8,9} because they allow short exposure times. For the ultimate resolution, ^3H or biotin-labelled probes are used. The unique emission characteristics of ^{125}I suggest an as yet largely unrealized role in high sensitivity, high resolution *in situ* hybridization. □

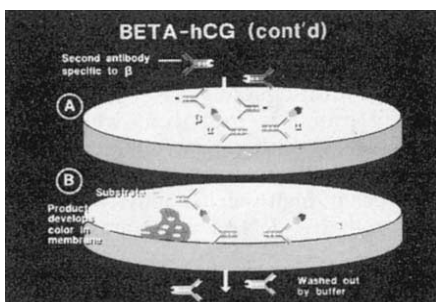
Martin W. Cunningham and Christopher R. Mundy are at Amersham International, White Lion Road, Amersham, Bucks HP7 9LL, UK. For a copy of Amersham's booklet entitled "Nucleic Acid Labelling" fill in Reader Service No. 100.

1. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
2. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. J. *molec. Biol.* **113**, 237-251 (1977).
3. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
4. Krieg, P. & Melton, D. *Nucleic Acids Res.* **12**, 7057-7071 (1984).
5. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
6. Cox, K. H., DeLeon, D. V., Angerer, L. M., & Angerer, R. C. *Dev. Biol.* **101**, 458-502 (1984).
7. Renz, M. & Kurz, C. *Nucleic Acids Res.* **12**, 3435 (1984).
8. Harper, M. E., Marselle, L. M., Gallo, R. C., Wong-Staal, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 772-776 (1986).
9. Lawrence, J. D. & Singer, R. H. *Nucleic Acids Res.* **13**, 1777-1799 (1985).

Putting labels into focus

The curtain lifts this week on tapes to teach immunoassays, and a cast of labelling reagents plus the instruments to detect them.

THERE is no Academy Award category for educational videotapes, but if there were, Savant Audiovisuals, Inc. would be in the running (Reader Service No. 101). Savant offers a range of programmes for **teaching immunoassay techniques**, and the newest is "Enzyme Immunoassays II: Practice", written by William F. Ulrich. The video, the sequel to "Enzyme Immunoassays I: Principles", discusses the interrelationships between enzyme and conjugate and

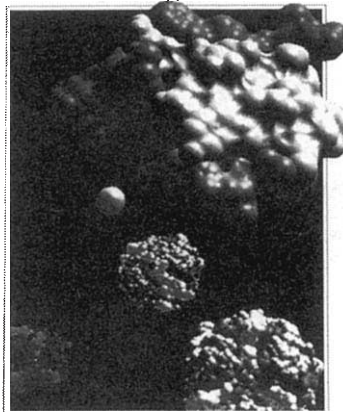


A sample still from the Savant videotape: an ideal way to teach immunoassay techniques.

the important parameters of sensitivity, precision, specificity and accuracy. Both homogeneous and heterogeneous assays (including ELISA) are described in detail, with an overview of current applications. The conclusion focuses on dose-response curves and quantitation. The 50-minute video programme costs \$540 (US), and is also available in slide/tape format for \$265 (US). Savant has a 15-day trial period for all educational materials.

The Calbiochem Biochemical/Immunochemical Catalogue for 1987 tells the full

Calbiochem
Biochemical/Immunochemical
Catalog 1987



Catalogue or reference manual? With 339 pages and a wealth of references, its hard to decide.

story on **biochemicals and immunochemicals** from Behring Diagnostics (Reader Service No. 102). The free 339-page book contains so much technical reference material, including bibliographies, that the company calls it a catalogue/reference manual. An applications-oriented presentation of the Calbiochem product line, the book includes chapters on immunization and antibody screening, agaroses for immunoprecipitation, immunoabsorbents, protein and nucleic acid modification reagents, RIA, EIA, IFA, antisera, and immunoglobulins and complement.

Dedicated detectives

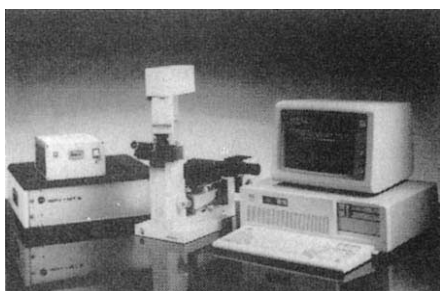
The Taurus automatic **liquid scintillation counter** from ICN Micromedex Systems, Inc. counts four standard or four mini-scintillation vials simultaneously (Reader Service No. 103). With preset isotope set-



The Taurus is a versatile workhorse of a liquid scintillation counter.

tings for ^3H , ^{14}C or ^{32}P , the \$39,900 (US) Taurus has built-in multiple assay stacking capability that allows the user to set up as many as six different assays to be counted using any of the 16 different protocols. Counts may be performed with or without external standard ratio quench correction, according to the user's needs, yielding either unquenched c.p.m. (raw data) or unquenched/quenched d.p.m. based on a single standard.

For semiautomatic **photometric evaluation of immunological reactions**, E. Leitz Instruments Ltd sells its MPV-MT2 photometric system (Reader Service No. 104). The Leitz system can be used for either incident light fluorescence or transmitted light measurements, and has a



The Leitz photometric system helps researchers keep an eye on immunological reactions.

scanning stage that accommodates all conventional microtest trays as well as slides of up to 76×26 mm. Suitable for a wide variety of applications, the MPV-MT2 can be used in virology, bacteriology, antibody screening, immunohaematology, and the interpretation of immunofluorescence and ELISA tests. The stage can be programmed for continuous or stop scanning, and two wavelengths for fluorescence excitation can be selected and changed during a sequence of readings by the motorized fluorescence illuminator.

The generation of photo-colloids for use in photo-affinity labelling can be accomplished with an instrument offered by Raytest Instruments (Reader Service No. 105). Known as Flash, the instrument uses flashlight photolysis techniques, producing pulses of ultraviolet light with an intensity range of $500\text{--}2,000 \text{ W s}^{-1}$ and a duration of 300 ms. To use Flash, samples are placed in two U-shaped quartz carrier tubes with a maximum volume of 10 ml, and exposed to the ultraviolet source in an enclosed box. Raytest says Flash can achieve 70 per cent disintegration of cyclosporin with a single ultraviolet pulse, as compared to conventional generation techniques that often require exposures of 20 minutes or more.

Fuji Film has developed a system for

film-less autoradiography that drastically reduces the developing time for autoradiographs (Reader Service No. 106). The Fujix BA-100 system uses a flexible polyester imaging plate coated with photostimulable phosphor to store radiation energy emitted from the sample. The plate is then 'developed' by exposure to a laser, and the resulting luminescence is converted into digital signals, which are in turn sent to a computer for storage and analysis. Using a special light source, the plate can then be erased for reuse. Autoradiographs can be performed in 20 minutes with the Fujix system, and because the sensitivity of the phosphor-coated plate is linear, accurate quantification of labelled areas is now possible. The Fujix system carries a price tag commensurate with its state-of-the-art sophistication, and currently sells for £120,000 (UK).

Tagging along with labels

Strattech Scientific Ltd has the solution for use in double labelling studies (Reader Service No. 107). Simultaneous detection of more than one antigen requires secondary antibodies that do not recognize each other, do not recognize each other's antigens (primary antibodies), and do not cross-react with conserved amino acid sequences. The probes must also be visually well-resolved. Jackson Affinipure double-labelling antibodies have been designed to meet these criteria. The antibodies are made in the same host species, and are available in fluorescence and enzyme immunodetection protocols.

Beckman is making scintillation counting less hazardous with the advent of its Ready Safe biodegradable scintillation cocktail (Reader Service No. 108). Ready Safe has a high flashpoint (150°C) to reduce fire hazards, and a low vapour pressure to prevent release of toxic fumes into



When through with the Beckman Ready Safe scintillation cocktail, just pour it down the drain.

the laboratory. Because it is biodegradable, it can be dumped down the drain if state and local laws and institution regulations permit. Beckman's cocktail does not sacrifice counting efficiency for safety, and Beckman claims it has a 30 per cent larger sample holding capacity than other

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



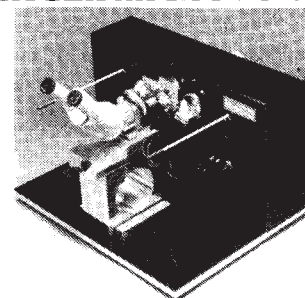
Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

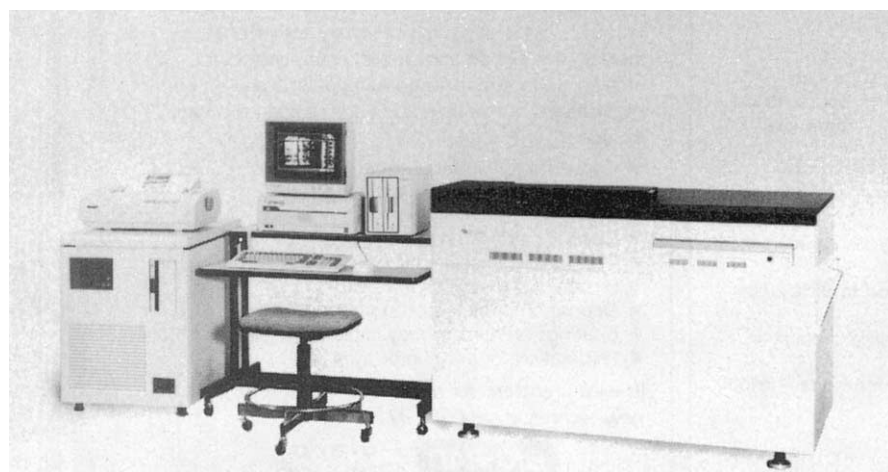
Reader Service No.35

SINGER INSTRUMENTS
SINGER INSTRUMENTS
MICRO-TOOL MAKING?
then you need the
SINGER MICROFORGE



Write to: Singer Instrument Co Ltd, Treborough Lodge, Roadwater, Watchet, Somerset TA23 0QL, England. Tel: (0984) 40226. Telex: 337492 Comcab G.

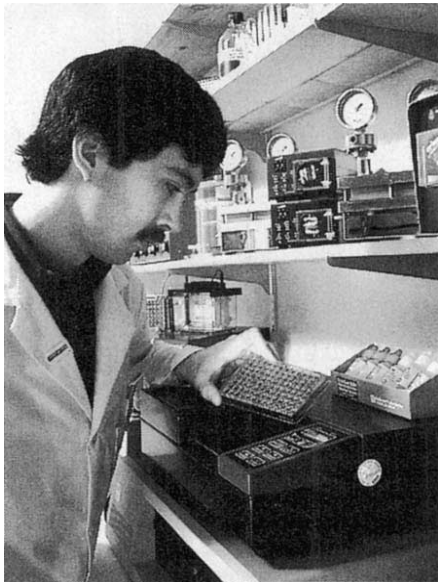
Reader Service No.50



Autoradiography without film! The Fujix system uses an image plate that can be erased and reused.

aqueous cocktails. Packaged in 5-litre metal safety containers, Ready Safe costs \$225 (US) for three containers.

BioGenex Laboratories is offering kits for the staining of enzyme-linked immunosorbent assays (ELISAs), immunoblots (Westerns), and large volume tissue staining (Reader Service No. 109). Based on a



BioGenex marries biotin-streptavidin to immunoreagents in its kits.

simplified biotin-streptavidin methodology comprised of a biotinylated anti-immunoglobulin (anti-mouse, anti-rabbit, or anti-goat) and streptavidin-conjugated enzyme (alkaline phosphatase, β -galactosidase, horseradish peroxidase), the \$140 (US) kits contain enough reagent to make 250 ml of each solution. Protocols detailing hybridoma screening, development of sandwich assays for antigen detection, and the staining of immunoblots are included in the kits.

Vector Laboratories has a myriad of labelling products for detection by flow cytometry or fluorescence microscopy, taking advantage of the intense fluorescent properties of R-phycoerythrin (Reader Service No. 110). R-phycoerythrin is derived from the light-harvesting system in red algae, and has an absorbance spectrum of 450-570 nm, and an emission maximum at 574 nm. Vector's array of R-phycoerythrin products includes R-phycoerythrin labelled anti-mouse IgG, avidin D, and Vectastain ABC-PE, a system that consists of many biotinylated R-phycoerythrin molecules crosslinked by avidin into a three-dimensional complex. The reagents range in price from \$75 (US) and \$65 for the anti-mouse IgG and avidin D, to \$105 (US) for the Vectastain ABC-PE.

Janssen has exploited the natural affinity of Protein A for IgG, and is selling two kits designed to help researchers in immunochemistry, immunocytochemistry, and immunoelectron microscopy reap the benefits (Reader Service No. 111). Janssen's £55 (UK) AuroProbe BL Plus



Janssen's AuroProbe BL Plus kit contains Protein A plus gelatin to block nonspecific binding. Protein A kit is suitable for immunoblot and dot blot applications. Each kit includes 2ml of Protein A and 5g of gelatin to block non-specific binding. Janssen recommends the use of its new IntenSE II silver enhancement reagents in liquid form with the AuroProbe kit for best results. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS



ROTEM INDUSTRIES OFFERS "One-of-a-Kind" TRITIUM LABELLING SERVICE

For:

- ★ Biology ★ Biochemistry ★ Pharmacology
- ★ Pharmaceuticals ★ Related life sciences

THREE PROCESSES FOR THE PRICE OF ONE

- ★ Labelling ★ Purification ★ Analytical control
- Including determination of the specific activity of the purified compounds

ROTEM LABELLING METHODS

- ★ Catalytic reduction with tritium gas
- ★ Catalytic dehalogenation with tritium gas
- ★ Catalytic reduction with tritiated sodium borohydride
- ★ Catalytic exchange in solution with tritium gas

MORE EXTENSIVE - LESS EXPENSIVE

The basic price for Rotem's labelling service covers:

- Reaction of the precursor compound
- If necessary, a one-step preparation of a brominated precursor
- Removal of the catalyst and labile tritium
- Purification of the labelled compound, guaranteed to 97% purity
- Chromatographic analysis of purity
- Measurement of the specific activity of the purified product
- Packaging of the final product

One price for any of the above labelling methods - \$1500

Our address:

Rotem Industries Ltd.
Radiochemistry Department
P.O.Box 9046 Beer-Sheva 84190 Israel

Phone: (57)-967755
Telex: 5375 RTMKG IL
Fax: 972-57-54502

PLASMA THE LIFEBLOOD OF PLASTICS ENGINEERING

Polaron's range of plasma chemistry equipment includes barrel type and parallel plate reactors for etching, surface modification and plasma deposition. Applications of this new and wide ranging technique include:-

- Conversion of surfaces to be hydrophilic
- Deposition of hydrophobic materials
- Deposition of low friction coatings
- Optical coatings on glass and plastic lenses
- Variable permeability plastic coatings
- Preparation of reinforcing fibres
- Deposition of metals onto plastics
- Coatings on human body implants
- Preparation for printing or dyeing

The list is endless, for more details on this exciting new technique call 0923 37144.

BIO-RAD Polaron Division

Polaron Equipment Limited

Greenhill Crescent, Watford Business Park,

Watford WD1 8QS England.

Tel: 0923 37144 Telex: 922080